TARGET LOCALIZATION AND TRACKING IN MEDICAL IMAGES

DESIGN OF A SYSTEM FOR TARGET LOCALIZATION AND TRACKING IN IMAGE-GUIDED RADIATION THERAPY

By

OLESYA PESHKO, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2016) McMaster University
(Computational Science and Engineering) Hamilton, Ontario

TITLE: Design of a system for target localization and tracking in image-guided radiation therapy

AUTHOR: Olesya Peshko, B.Sc., M.Sc.

SUPERVISORS: Dr. Timothy N. Davidson, Dr. Jan Modersitzki, Dr. Douglas J. Moseley, and Dr. Tamás Terlaky

NUMBER OF PAGES: xxxi, 377

To my family:
Igor, Nataliya, and Bogdan Peshko,
and Bogdan Kruts

Lay Abstract

This thesis presents the development of a software system that analyzes sequences of 2D x-ray images to automatically measure organ motion in patients undergoing radiation therapy for cancer treatment. The knowledge of motion statistics obtained from this system creates opportunities for patient-specific treatment design that may lead to a better outcome.

Automated processing of organ motion is challenging due to the low contrast and high noise levels in the x-ray images. To achieve reliable detection, the proposed system was designed to make use of 3D cone-beam computed tomography images of the patient, where the features (markers) are easier to identify. This required the development of a specific image registration framework for aligning the images, including a number of novel feature modelling and image processing techniques.

The proposed motion tracking approach was implemented as a complete software system that was extensively validated on phantom and patient studies. It achieved a level of accuracy and reliability that is suitable for clinical implementation.

Abstract

This thesis contributes to the topic of image-based feature localization and tracking in fluoroscopic (2D x-ray) image sequences. Such tracking is needed to automatically measure organ motion in cancer patients treated with radiation therapy. While the use of 3D cone-beam computed tomography (CBCT) images is a standard clinical practice for verifying the agreement of the patient's position to a plan, it is done before the treatment procedure. Hence, measurement of the motion during the procedure could improve plan design and the accuracy of treatment delivery. Using an existing CBCT imaging system is one way of collecting fluoroscopic sequences for such analysis. Since x-ray images of soft tissues are typically characterized with low contrast and high noise, radio-opaque fiducial markers are often inserted in or around the target. This thesis describes techniques that comprise a complete system for automated detection and tracking of the markers in fluoroscopic image sequences.

One of the cornerstone design ideas in this thesis is the use of the 3D CBCT image of the patient, from which the markers can be extracted more easily, to initialize the tracking in the fluoroscopic image sequences. To do this, a specific marker-based image registration framework was proposed. It includes multiple novel techniques, such as marker segmentation and modelling, the marker enhancement filter, and marker-specific template image generation approaches.

Through extensive experiments on testing data sets, these novel techniques were combined with appropriate state-of-the-art methods to produce a sleek, computationally efficient, fully automated system that achieved reliable marker localization and tracking. The accuracy of the system is sufficient for clinical implementation. The thesis demonstrates an application of the system to the images of prostate cancer patients, and includes examples of statistical analysis of organ motion that can be used to improve treatment planning.

Acknowledgments

First and foremost, I would like to thank my supervisors: Dr. Timothy N. Davidson, Dr. Tamás Terlaky, Dr. Douglas J. Moseley, and Dr. Jan Modersitzki. The opportunity to work with these four brilliant people pushed me forward and motivated me to grow.

Being an accomplished researcher and talented administrator, Dr. Tamás Terlaky created an environment that gave me the opportunity to complement my classroom learning of optimization and image processing with numerous attendances and presentations at seminars and conferences. My transformation into an engineer and a scientist who can design and implement a challenging project from beginning to the end was made possible thanks to his continuous support.

I learned a lot about radiation therapy and medical imaging from Dr. Douglas J. Moseley, who helped me formulate initial requirements for the problems discussed in this thesis. He was my guide in peculiarities of the field, in design of experiments and data collection, as well as in data analysis and validation procedures. Dr. Moseley helped me understand and appreciate the difference between models and real systems, and a fine balance between simplicity, reliability, and accuracy.

In addition to specific knowledge of image registration topics, Dr. Jan Modersitzki taught me to ask questions, think critically, support my claims, formulate problems, and use both mathematical and linguistic formulations responsibly. Discussions with him brought me to an understanding of how many things in mathematics (and the world which mathematics tries to model) are interconnected, and how the same subject can be examined from various viewpoints.

The guidance of Dr. Timothy N. Davidson ranged from help with digital filtering and optimization topics to extensive aid with the writing process. While multiple valuable comments and corrections were provided by all other PhD supervisory committee members, Dr. Davidson was there for me throughout the whole process of thesis writing: from the very first drafts to the finished work. I would like to express special gratitude to him for acting as a balancing and unifying force between all committee members, and taking care of administrative issues arising during my studies.

I would like to deeply thank all my supervisors for financial support, as well as McMaster University, Mathematics of Information Technology and Complex Systems (MITACS), Princess Margaret Hospital (PMH, University Health Network, ON), and Elekta Oncology Systems (Atlanta, GA) for scholarships, internships, and conference funding.

I would like to express my great appreciation to Dr. Aleksandar Jeremic and Dr. Michael B. Sharpe who served on my PhD supervisory committee. They provided valuable comments and discussions, and encouraged me through the difficulties.

Special thanks go to Dr. Cynthia Ménard and Dr. Tim Craig from PMH for providing patient images, without which the comprehensive testing of the feature localization and tracking system proposed in this thesis would not be possible. I would like to express gratitude to Dr. David Jaffray who helped organize my collaboration with PMH, as well as Dr. Yuriy Zinchenko who encouraged me to start the PhD program and made this thesis project possible by initial introductions to some of the aforementioned PMH researchers.

I greatly appreciate administrative help provided by Dr. Antoine Deza from Computing and Software department, McMaster University, as well as Laurie Leblanc, David Ryan, Laura Kobayashi, and Tina Macala.

I am indebted to express the warmest thanks to my family. I would like to thank my parents, Dr. Igor Peshko and Nataliya Peshko, for their endless love, support, and an always welcoming attitude. By example and encouragement, they played an important role in developing my interests in study, research, and fine arts. My brother, Bogdan Peshko, is in charge of a lot of fun in my life. Being a good role model for an active lifestyle, he encourages me to explore the world and try new things. He is my source of the most stimulating discussions and enormous sibling support. I would like to thank my partner, Bogdan Kruts, for making me happy and supporting me through the challenges in so many ways. His love and aptitude for self-development encourages me in my search of a better self.

Finally, I would like to thank my friends and colleagues from McMaster University. Dr. Dmitry Labukhin supported me in one of my toughest times and encouraged me to start writing this thesis. He was the one to read and comment on the very first pages. I appreciate friendship, help, support, and advice provided by Nadezhda Zubkova, Shefali Kulkarni-Thaker, Drs. Kimia Ghobady, Imre Pólik, Vera

Pantelic, Bahareh Mansouri, Nick Kisialiou, Ramy H. Gohary, Sergiy A. Vorobyov, Anton Gorny, and Oleksandr Romanko. Special thanks to Dr. Vadim Aksenov for his support, friendship, and for helping me on so many occasions. I would also like to acknowledge all members of the Advanced Optimization Laboratory. Finally, I would like to thank all my friends for making my life interesting by creating opportunities and challenges outside of university.

Contents

List of Flowcharts

List of Figures

List of Tables

List of Algorithms

Abbreviations

2D, 3D, 4D	two-, three-, and four-dimensional
AP	anterior-posterior
CBCT	cone-beam computed tomography
CEF	contrast enhancement function
CNR	contrast-to-noise ratio
CT	computed tomography
CTV	clinical target volume
DRR	digitally reconstructed radiograph
EPID	electronic portal imaging device
fMRI	functional magnetic resonance imaging
GTV	gross tumour volume
IGRT	image-guided radiation therapy
IMD	inter-marker distance
ITV	internal target volume
LA	linear adaptive
LE	linear extrapolation
LINAC	linear accelerator
LR	left-right, linear regression
MEF	marker enhancement filter
MIP	maximum intensity projection
MM	marker model
MOF	magnitude-only filter
MLC	multileaf collimator
MRI	magnetic resonance imaging
MS	marker splat
NCC	normalized correlation coefficient, normalized cross-correlation
OAR	organ at risk
PET	positron emission tomography
PTV	planning target volume
ROI	region of interest
RSVP	radio-surgery verification phantom
RT	radiation therapy
SI	superior-inferior

SPECT	single-photon emission computed tomography
VOI	volume of interest
VS	voxel splat
XVI	X-ray Volume Imaging (a software by Elekta)
ZO	zeroth-order

The following abbreviations pertaining to measurement units are used in this thesis:

Hz	Hertz, number of frames per second in image acquisition
kV	kilovolts, kilovoltage
kVp	peak kilovoltage
mA	milliamperes
mm	millimeters
ms	milliseconds
MV	megavolts, megavoltage
px	pixel, pixels
sec	seconds
vx	voxel, voxels

Notation and Symbols

General Assumptions

Symbols for mathematical formulations used in this thesis generally adhere to the following principles:

Alphabet, font	Notation for	Examples
Calligraphic capital	functions and operators; continuous images	$\mathcal{D}$, $\mathcal{P}$; $\mathcal{V}$, $\mathcal{M}$
Teletype font	discrete images	$\mathtt{S}$, $\mathtt{R}$
Greek capital	sets	Ω, Γ
Greek and Latin lower case italic	parameters	ϱ, r
Latin capital italic	matrices	A, F
Latin lower case upright	functions	diam

Numbering of vectors, parameters, sets, and images has the following convention:

- Symbols i, j are used for counters, and d to denote a space dimension.
- The notation for vectors $p \in \mathbb{R}^d$ always assumes that p is a column-vector, where the elements of p are denoted as $p = (p_i) = (p_1, \ldots, p_d)^T$. The notation $p^1, \ldots, p^n$ refers to n vectors, of the same dimension.
- The numbering of parameters uses subscripts, e.g., λ_1.
- Images and sets are also numbered with the subscripts, e.g., $\mathcal{S}_0$, M_0.

Unless specified otherwise, the norm of the vector $\|p\|$ refers to the L_2-norm, i.e., $\|p\|_2 = \sqrt{p_i^2 + \ldots + p_d^2}$.

In the remainder of this section, we provide lists of symbols used globally throughout this thesis.

Images

Discrete images (original data):

$\mathtt{V}$	3D cone-beam computed tomography (CBCT) image
$(\mathtt{S}_i)$	2D fluoroscopic image sequence

Continuous images (image models):

$\mathcal{V}$	3D CBCT image model
$(\mathcal{S}_i)$	2D fluoroscopic image sequence model

Preprocessed and digitally generated:

$(\bar{\mathcal{S}}_j)$	temporally filtered 2D image sequence
$\mathcal{R}$	MEF-image (result of an application of the marker enhancement filter)
$\mathcal{M}$	selective digitally reconstructed radiograph (DRR)

Sets

Γ^d	d-dimensional grid on which a discrete image is defined
M_i	marker sets

Geometric Transformations

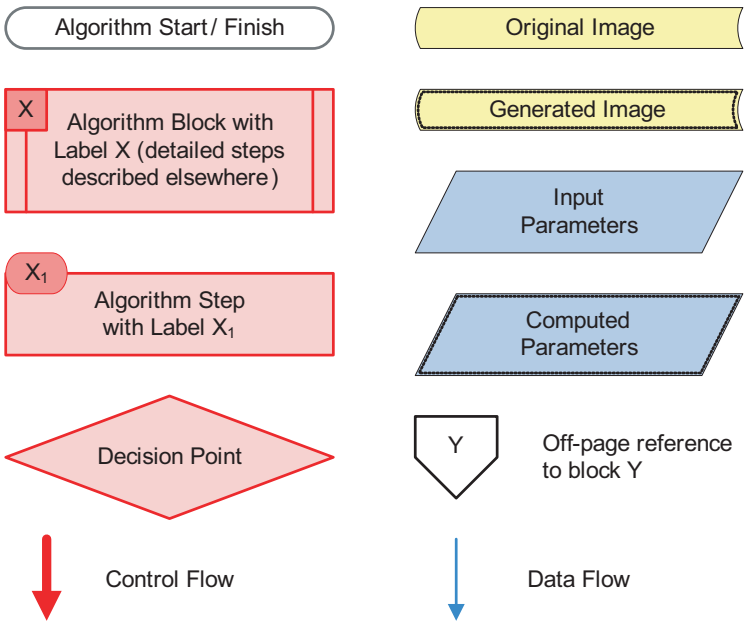
$\mathcal{T}^{3\mathrm{D}}(a, p)$	3D rigid transformation of point $p \in \mathbb{R}^3$ parameterized by $a \in \mathbb{R}^6$ (three translations and three rotations)
$\mathcal{T}^{2\mathrm{D}}(b, p)$	2D rigid transformation of point $p \in \mathbb{R}^2$ parameterized by $b \in \mathbb{R}^3$ (two translations and a rotation)
$R^{3\mathrm{D}}$, $R^{2\mathrm{D}}$	3D and 2D rotational matrices
$T^{3\mathrm{D}}$, $T^{2\mathrm{D}}$	3D and 2D translation matrices
$\mathcal{P}$	perspective transformation from XYZ to UV coordinate systems

Other Functions

$\mathcal{D}$	distance measure between two images
diam	marker set diameter
dst	the shortest distance between the sets
min, max	minimum, maximum
argmin	argument for which the minimum function value is achieved

Flowchart Notation

In this thesis, flowcharts will be used to capture, graphically, the interactions between components of the proposed marker localization and tracking system. To simplify the reading of these flowcharts, different components will be represented in different styles. The following diagram explains the shapes that are assigned to the different components.



Chapter 1

Introduction

Radiation therapy is a powerful tool in treatment and control of cancer. It is prescribed to nearly half of all cancer patients in Western countries [77]. *Image-guided radiation therapy* is a widely accepted technique to visualize inner structures in a patient's body to verify the position of the target immediately before or during a radiation treatment procedure. The imaging is typically performed with a diagnostic imaging x-ray, e.g., [243, 152, 159, 205]. However, the desire to reduce the imaging x-ray dose to the patient often limits the use of image guidance to initial setup position verification. This verification is necessary to confirm that the actual patient position at the beginning of the treatment conforms to the treatment plan. While this helps to adjust the position and compensate for possible anatomical changes, it does not account for all positioning errors introduced during treatment preparation and execution [219]. In addition, since the imaging is performed before the treatment procedure, the setup verification is inherently unable to deal with an internal organ motion that happens *during* treatment delivery, referred to as an *intra-fraction motion*. To ensure that the treatment target in a patient's body actually receives the planned dose of radiation in spite of the motion, the target volume is extended by a *safety margin* [219] that is supposed to cover most target displacements. Ideally, the size of the margin reflects assumptions on magnitude of target positioning errors and organ motion.

The ability to automatically measure the intra-fraction motion has the potential to improve margin design by making it possible to incorporate realistic motion statistics for each patient or certain patient cohorts. In particular, retrospective statistics or real-time motion measurements may indicate the possibility to *reduce* the currently accepted margin size for certain anatomical sites, e.g., [144]. The goal of the margin reduction is to achieve the same level of tumour control, while aiming for better long-term health outcomes for the patients by decreasing the damage done by radiation to the healthy tissues and organs surrounding the tumour [67, 191]. Reducing the

margin also has the potential to enable a safe *dose escalation* [21, 67], a technique of increasing the radiation dose delivered to the target in one treatment fraction for improved treatment result and potential reduction in the number of fractions [208]. This, in turn, may allow to increase the number of patients that a clinic can accept.

This thesis presents an approach to the automatic quantification of intra-fraction motion from *fluoroscopic image sequences*, which are temporal sequences of two-dimensional (2D) kilovoltage x-ray images [198, 214]. In brief, the main challenges for automatic motion analysis are the *noise* and the *low contrast* in fluoroscopic images. Although it is possible to improve acquisition image quality by increasing the imaging x-ray dose, the dose management philosophy currently adopted by the diagnostic imaging community is "as low as reasonably achievable" [158], which is due to the desire to reduce possible radiation-related complications to the patient's health. Therefore, to achieve reliable image analysis, the difficulties associated with the noise and low contrast should be overcome by the development of the appropriate digital processing methods.

The culmination of the technical contributions of this thesis is the development of a software system, named Gryphon[1], for the automated processing of multiple image sequences and collecting of the intra-fraction organ motion statistics. Using the innovative techniques developed in this thesis, Gryphon performs an *automatic localization*, i.e., it determines the position of the target in the first image of each sequence. Then, it continues to *motion tracking* of the target by analyzing the subsequent images of the sequence. In the end, the motion statistics are reported back to the user. On a very high level, the heart of the localization and tracking procedures consists of image processing and image registration techniques. It also includes the use of a priori information, such as the details of the geometric setup and incorporation of other images of a patient, so that the system is able to achieve reliable results in a computationally efficient fashion.

To help place the technical discussion of the following chapters in context, this introductory chapter begins with the discussion of the problem of organ motion in radiation therapy in Section 1.1, including details on the use of imaging in radiation treatment procedures, hardware and software, positioning uncertainties caused by organ motion, and a review of current advances in motion management. While these topics are well covered in the domain specific literature, the inter-disciplinary nature of this thesis calls for an appropriate introduction to the subject matter. Section 1.2 provides an explanation of the goals and challenges of the work reported in this thesis,

[1] In antique times, as described in myths and legends of multiple nations, gryphons were thought of as magical creatures that guard hidden treasures, or, more generally, sacred knowledge. It seemed to be an appropriate title for a system intended for medical image processing and detection of features in low-contrast images.

a description of the problem solving strategies used, and a brief presentation of the main contributions. The chapter is concluded by the thesis outline in Section 1.3.

1.1 Problem of Organ Motion in Radiation Therapy

1.1.1 Use of Imaging in Radiation Therapy Procedures

Radiation therapy (RT) is the medical use of ionizing radiation as part of cancer treatment. Its therapeutic effect results from the fact that malignant cells have lower recovery rates than normal cells following the deoxyribonucleic acid (DNA) damage that is caused by the ionizing radiation. In this thesis, by RT we mean an *external-beam conformal therapy*, that is, a treatment that is designed in such a way that high-energy radiation from an *external source* is delivered to a target. The *target*, or *target volume*, closely conforms to the region in a patient's body that is supposed to receive the prescribed amount of radiation according to the treatment plan [62]. Anticipating a terminology that will be defined later, unless otherwise stated, by target volume we mean the planning target volume (PTV, Section 1.1.4).

The approach to RT treatment planning can be described in terms of risks and benefits [246]. Ideally, most malignant cells will be destroyed by being exposed to sufficiently high doses of radiation. However, in general it is not possible to deliver this radiation to the tumor without affecting surrounding healthy tissues. Therefore, the main task of radiation therapy is to design and deliver such a treatment regimen that provides the maximal possible tumor control and, at the same time, does not produce unacceptable damage to the healthy tissues.

Figure 1.1 shows a hardware setup for RT delivery on a *medical linear accelerator* [99] (LINAC, Elekta Oncology Systems, Atlanta, GA) with an integrated imaging system. A high energy radiation source is located in the head of the machine (C), which is capable of rotating around a patient lying on a couch to allow a choice of the best angle(s) for radiation entry. Typically, the patient is positioned so that a treatment target appears in the centre of the machine's rotation, called the *isocentre*. The system depicted in Figure 1.1 allows acquisition of x-ray images of the patient in the treatment position produced by using the kilovoltage x-ray source (A) and a flat-panel detector (B, hidden behind the patient), both of which rotate together with the treatment head (C).

To conform to the target, the radiation beam must be shaped. A *multileaf collimator* (MLC) [99], shown in Figure 1.2, is a device made up of individual slices, or leaves, of a material with a high atomic number, such as tungsten. The leaves

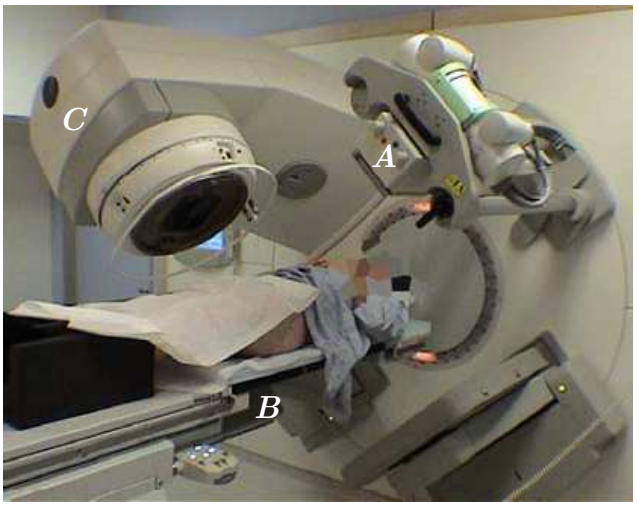


Figure 1.1: A patient on a medical linear accelerator equipped with an imaging kilovoltage x-ray source (A) and a flat-panel detector (B). A multileaf collimator shown in Figure 1.2 is located inside the linear accelerator treatment head (C), in the path of the megavoltage beam.

Figure 1.2: A multileaf collimator shapes the radiation beam to conform to the target in the patient's body [2].

can move independently in and out of the path of the beam to block a part of it. The MLC is installed in the LINAC's treatment head (Figure 1.1, C). Although the LINAC creates an essentially uniform dose distribution, the intensity of the radiation beam can be modulated by varying positions of the leaves in the MLC during the treatment delivery. This technique is known as an *intensity-modulated radiation therapy* (IMRT) [156], an extension of conformal therapy that is widely utilized today.

The choice of the position for the patient's treatment, as well as the necessity to use immobilization devices, depends on several factors, the most important of which are *patient comfort and stability*, and the *beam directions* (angles) that will be used [62], which, in turn, influence the effectiveness of the treatment. In order to reduce the radiation impact on the healthy tissues adjacent to the tumor, conformal therapy typically uses three or more radiation beams directed from different positions intersecting in the target volume [62], where the prescribed dose accumulates. Another technique to spare healthy tissues is *fractionation* [48]. It has been noted that healthy cells affected by radiation have a better chance of recovery than cancerous cells. Therefore, the prescribed dose is divided into several fractions that are delivered on different days to allow normal tissue renewal between the fractions. Often, treatment planning is only performed once, therefore, the *reproducibility of the patient position* for each treatment procedure is extremely important.

Conformal therapy process involves several steps:

1. **Treatment simulation.** A conformal therapy plan is based on a volumetric *anatomic model* of the patient that is derived from a three-dimensional (3D) or, sometimes, four-dimensional image (4D, a temporal sequence of 3D images) showing a target volume and adjacent critical structures. The critical structures are healthy tissues or organs, to which radiation damage should be avoided. They are often referred to as the *organs at risk* (OAR). Typically, the target and surrounding OARs are represented in the anatomic model as sets of contours or surfaces. In most of the cases, a *computed tomography* (CT) data set is used as it provides the necessary anatomic information, and represents a geometrically accurate base for planning [62]. However, it does not provide any physiologic or functional information, and is characterized by *limited soft-tissue visualization*, i.e., there is no or very little contrast between different types of soft tissues. To overcome this drawback, other techniques, such as magnetic resonance imaging (MRI), functional MRI (fMRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT), can be used in addition to CT (a detailed discussion on image acquisition methods in application to RT is presented by Evans [53]). To be utilized for treatment planning purposes, these additional data sets have to be geometrically aligned with the original CT scan.

2. **Treatment planning.** Radiation therapy planning is a computer simulation of the treatment process. It involves recreating the 3D anatomic model of the patient inside the planning software, simulating radiation beams and the dose delivered to the patient. This virtual model of the patient contains the target (or targets) and OARs in the proximity of the target. One of the common ways to describe these structures is through a series of contours on the CT slices. To date, this is one of the most time-consuming parts of the RT planning process due to the large amount of manual labour. Even with the use of image segmentation software, the contours may need to be verified and corrected manually, slice by slice, the total number of which may easily exceed 100 [62].

3. **Treatment delivery.** The delivery of the prescribed radiation dose to the target can not be achieved if the patient's position at each treatment fraction is not reproduced with sufficient accuracy. For many years, the initial setup of the patient on a medical LINAC was performed using conventional laser alignment in the room to patient's skin marks, and immobilization devices. Weekly portal films were obtained to verify the correctness of the target position, where the term *portal* refers to the 2D images produced with the help of the treatment megavoltage (MV) x-ray source on a film or a digital detector. In general, portal

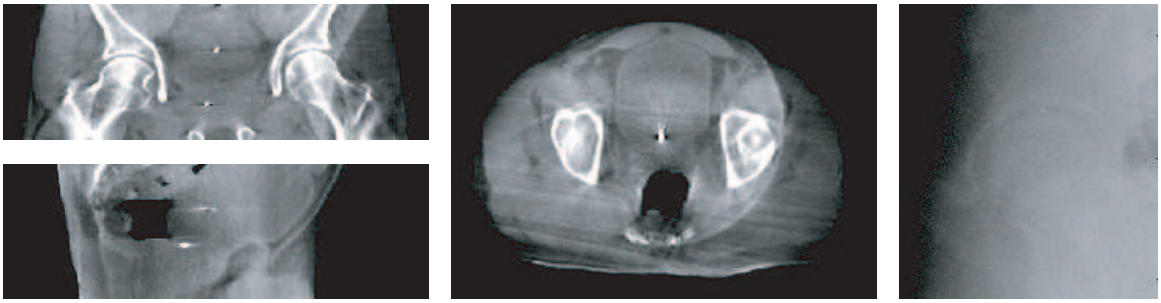


Figure 1.3: CBCT image slices of a pelvic area of a prostate cancer patient (*left and middle*) and a corresponding fluoroscopic/kV x-ray projection image (*right*).

images are characterized with lower signal quality than images obtained with a kilovoltage (kV) x-ray source. Nowadays, clinics perform high accuracy automated patient setup, enabled by such technologies as *electronic portal imaging devices* (EPIDs) [10] for MV imaging, *diagnostic kV x-ray imagers* [11], and *cone-beam computed tomography scanners* (CBCT) [92] mounted directly on the LINAC as depicted in Figure 1.1. Hence, a therapist not only reproduces the position of the patient based on the external skin marks and immobilization devices, but can actually verify the position of internal structures before each treatment fraction and make appropriate adjustments if necessary.

1.1.2 Imaging Modalities

In this thesis, we will use the following imaging modalities, demonstrated in Figure 1.3[2]:

- **X-ray projection** (radiography) is a single 2D projection image of an object. Structures contained in an imaged volume appear to be overlaid in the projection image. The projection can be produced on a detector by an MV or kV beam. To differentiate between the two, we will use the term x-ray projection to refer to the kV image, and portal or MV image to refer to the MV projection.
- **Fluoroscopic image sequence** is a series of kV x-ray projections characterized by some temporal resolution, for example, 5.5 to 30 Hz (frames per second), acquired with the stationary imaging system. We use fluoroscopic sequences to assess organ motion from a single viewpoint ("in-plane" motion, i.e., motion that can be observed in the direction parallel to the flat-panel detector). The term *fluoroscopic frame* refers to a single image within the sequence.

[2]In this thesis, images of the patients are courtesy of Dr. Cynthia Ménard, Dr. Timothy Craig, and Dr. Douglas Moseley, Princess Margaret Hospital, University Health Network, Toronto, Ontario.

- **CT image** and, likewise, **CBCT image** are both 3D images, where each slice can be viewed individually. Typically, the CBCT image is reconstructed by applying the *filtered back-projection algorithm* [94] to a set of 2D x-ray projections acquired by the imaging system from different viewpoints as the gantry of the LINAC rotates around the patient. CBCT technology allows reconstruction of an entire image from a single gantry rotation. It may take around 1–2 minutes to acquire and reconstruct. Naturally, the number of projections used in the reconstruction affects the reconstructed image quality as, roughly speaking, the angle intervals between the subsequent projections correspond to missing information resulting in shadows and artifacts in the final CBCT image. So far, CT images offer better quality than CBCT. However, they are acquired on a designated CT scanner. As a result, CT images can not be produced shortly prior to or during the treatment procedure with the patient in the treatment position.

1.1.3 Hardware and Software

Imaging of the patient before or during radiation treatment to direct the process of radiation delivery and verify its compliance with the treatment plan is often referred to as the *image-guided radiation therapy* (IGRT) [91, 243]. For treatment, the patient has to be placed in a way that reproduces their position during the planning CT acquisition as accurately as possible. Two-dimensional IGRT usually involves matching kV or MV x-ray projection images to the *digitally reconstructed radiographs* (DRRs). The DRRs are simulated 2D projection images computed from the 3D planning CT that imitate the geometrical setup and, if possible, intensity characteristics of the kV or MV x-ray projections. Three-dimensional IGRT compares the 3D CBCT image acquired immediately prior to treatment (daily CBCT) to the planning CT.

Naturally, for the success of the IGRT, it is strongly desirable that the hardware for image acquisition is located in the close proximity to the LINAC, so that the patient needs to not be repositioned between the daily target position verification imaging and the treatment delivery stages. Figure 1.1 demonstrates a patient resting on an Elekta medical LINAC with an integrated kV imaging system that is capable of radiography, fluoroscopy, and CBCT. The imaging kV x-ray tube (A) is located on a retractable arm at 90° to the MV treatment x-ray source (C). A 41×41 cm^2 flat-panel x-ray detector (B) is mounted opposite to the kV tube. The imaging system is supplied with a software package called XVI (X-ray Volume Imaging, Elekta Oncology Systems, Atlanta, GA) that is capable of calibration, image acquisition and processing, and CBCT reconstruction. Information about patients, treatments, and all images acquired under the control of XVI is stored in a database. The detailed technical

description of an IGRT system similar to the one used in this thesis is provided by Jaffray *et al.* [92].

1.1.4 Uncertainties in Target Position Due to Organ Motion

To fight the disease or achieve palliative effects, it is important to precisely specify the volume in the patient's body that is supposed to receive high-energy radiation treatment. The classification of the volumes in radiation oncology established by the International Commission on Radiation Units and Measurements (ICRU) is as follows [4, 6]:

- **Gross tumour volume** (GTV) is the volume of the macroscopic tumour, involved lymph nodes, and metastatic disease that is either palpable or visible on images.
- **Clinical target volume** (CTV) is the volume that should be treated to high dose, typically incorporating both the GTV and subclinical microscopic disease. The CTV is usually designed by expanding the GTV by a margin (typically, 0.5–1 cm [62]) that accounts for microscopic invasion. Additional regions can also be included in the CTV based on clinical experience with standard directions of disease spread for particular anatomical sites [62] and pathology studies [99]. In many cases, the CTV cannot be considered a stable volume, as it may vary in size, shape, and position due to physiological processes, such as filling of the bladder, respiration, etc. In this case, the **internal target volume** (ITV) is defined as the volume encompassing all known variations of the CTV.
- **Planning target volume** (PTV) is the volume that should be irradiated to ensure that the CTV (or ITV) receives the prescribed dose. It is designed with consideration of positioning uncertainties and organ motion. The PTV is often created by expanding the CTV borders isotropically in three dimensions. However, for some sites, it may be desirable to expand the margin in an anisotropic manner based on anatomy knowledge and the results of organ motion assessment for a certain patient or patient cohorts, e.g., [129, 147]. In the previous sections, by referring to the target volume we meant the PTV, and we will retain this terminology for the rest of this thesis, unless an explicit differentiation is required.

Expanding the PTV margin so that it always covers all variations of the CTV satisfies the goal of obtaining maximal tumour control. However, this tactic can cause unacceptable damage to surrounding healthy tissues and organs, the OARs, thus defeating the overall purpose of treating the patient. Naturally, it is desirable to keep the PTV margin at the necessary minimum.

In radiation therapy, *organ motion* is loosely defined as the displacement of the organ, or the target volume, in time, where the time interval can be as large as the break between two different treatments. Among different types of motion observed in patients during a course of radiation therapy, for the purpose of this thesis it is important to define the following two [116]:

- **Intra-fraction** motion is the motion within one fraction, such as respiration, heartbeat, peristalsis, swallowing, etc.
- **Inter-fraction** motion refers to the motion between the fractions, such as changes in the patient anatomy (weight gain/loss), CTV shape and position (often caused by daily changes in digestive system). To avoid confusion, we will use the term inter-fraction displacement rather than inter-fraction motion.

The ability to precisely quantify organ motion for a specific anatomical site of different patients, or even a particular patient, has direct implications for the PTV margin design. Different approaches to derive formulas for margin size computation based on measured motion have been proposed, e.g., [220, 219, 146] and references therein. In an effort to minimize the influence of inter-fraction displacement, many clinics have implemented daily CTV position verification and correction protocols by using imaging devices integrated into the treatment machines, e.g., [92, 152]. Another advancement is the clinical process termed *adaptive radiotherapy* (ART), in which the treatment plan is modified based on changes observed in daily images of the target, e.g., [144].

Management of intra-fraction motion ranges from dietary restrictions [8, 16, 206], to external body motion tracking [116, 230], to internal target motion measurement or monitoring [71, 230]. In fact, it can even start at the planning image acquisition stage: For some sites exhibiting large motions, such as the lung or liver, it is desirable to incorporate CT scanning protocols that do not add motion-related artifacts, e.g., *4D CT* [133] or *respiratory-correlated CBCT* [207]. During the treatment, the radiation dose to normal tissues can be reduced by such techniques as *respiration/cardiac gating* [50, 93, 111, 199, 229] or *tumour tracking* [157, 177, 194, 244]. Respiration gating allows radiation exposure only when the target is in the path of the beam, while tumour tracking methods aim to track the target and modify the beam to follow the target's trajectory dynamically. Both approaches need real-time information regarding the target's position.

One of the most straightforward approaches to control the respiration induced organ motion is breath-hold, either voluntary [100, 136] or controlled [82, 107, 210]. The *active breathing coordinator* (or active breathing control, ABC) [237] is a device that works in coordination with the radiation beam and induces repeat breath holds in patients at certain times. Another successfully utilized technique is *visual* or *audio-*

visual feedback [128], when the patient is looking at the trace of their breathing motion in time and is asked to comply to an appropriate pace and deepness of breath.

As not every patient can be put on the breath hold [65] or adapt their breathing to restrictions of audio-visual feedback, many clinics develop alternative methods for target tracking, for example, *abdominal surface motion monitoring* [61, 226, 222]. In this technique, the external marker is incorporated into a block that is placed on the patient body surface and is tracked with the infrared camera. Other approaches include the measurement of respiratory volume using a *spirometer* [111, 247], the use of a *thermometer* placed in the breathing airflow in the patient's nostrils [236], and a *strain gauge*, in the form of an elastic band that is wrapped around the patient's torso [111]. However, studies showed that the correlation between the motion of the external surface and internal target can be weak and non-reproducible [87]. Due to this fact, and also because the patient exhibits other types of motion that can not be measured externally, such as peristaltic motion, a natural advancement in tracking techniques involves assessing trajectories of internal structures. One of the approaches, fluoroscopic tracking, is reviewed in the next section.

1.1.5 Fluoroscopic Tracking of Organ Motion

Due to relative affordability of kV x-ray imaging systems, and the technical ability to mount them on or in the close proximity to the LINAC, *fluoroscopic tracking* is an area of active exploration. However, in addition to exposing a patient to an imaging radiation dose, the x-ray technology has another major drawback in the context of internal organ motion tracking. As soft tissues are characterized by low contrast in x-ray images, it may be very challenging to detect the target in many anatomical sites, especially automatically. One of the approaches to bypass this problem is to perform fluoroscopic tracking relative to the high-contrast *bony anatomy*. However, it was shown that the correlation between the motion of the bones and the target is not always strong enough for the bones to be used as a reliable position identifier for the target [74, 161, 190]. Another research direction addresses fluoroscopic *tumour mass tracking* through image intensity change [24, 50] and by using machine learning techniques [49, 124]. This is typically applied to lung cancer studies.

Taking into account aforementioned low contrast of soft tissues, the position of the target can be obtained more easily by tracking the *fiducial markers*, inserted in the proximity to the target, locations of which have been shown to be accurate *internal surrogates* for the position of the target [103, 198, 217]. Note that the use of the markers may be inadvisable for some sites, such as the lung, due to possible clinical complications of percutaneous intervention [14, 66]. For other sites, such as the prostate, the use of the fiducial markers is an accepted clinical practice [152].

Currently, there are two types of internal markers used: metallic radio-opaque markers [152], measurements of which are visible in the x-ray and CT/CBCT images, and electromagnetic transponders [113]. The electromagnetic transponders provide reasonably reliable target position information without additional x-ray exposure [113] but the use of the transponders alone does not produce images, which is often desired. Also, these markers are larger in size[3] and more expensive than the typical radio-opaque markers. To reduce an additional dose to the patient caused by the use of the radio-opaque markers and x-ray or CBCT imaging, the methods combining the external motion tracking with occasional x-ray imaging were developed [193, 95, 44, 240].

This thesis presents a detailed description of a system for automatic quantification of intra-fraction organ motion from fluoroscopic image sequences that is based on fiducial marker tracking.

1.2 Goals, Approaches, Challenges, and Results

The main goal of this thesis is to propose a set of techniques that constitute a complete system for automatic post-processing of multiple fluoroscopic image sequences in order to measure patient intra-fraction organ motion. To validate these techniques, and to demonstrate the collection of motion statistics for prostate cancer patients, a software system called Gryphon was developed. The following chapters present problem requirements and assumptions, mathematical models, algorithmic details, parameter selection approaches and system validation in terms of accuracy and reliability.

To infer the displacement of the target that occurred between the acquisitions of two consecutive frames of a fluoroscopic sequence, *image registration* [149, 150] techniques are used. Image registration is the process of finding a geometrical transformation between the objects portrayed in the images by computing the best alignment between the images. In order to be computationally efficient and reliable, i.e., legitimate and meaningful for the proposed application, it is desired that the registration be performed over small *regions of interest* (ROIs) in the images that include measurements of the target. *Automatic selection* of such regions constitutes the main challenge. According to the clinical protocol, in order to provide an easier target identification, all patients whose images we analyze had three small *radio-opaque fiducial markers* inserted in the proximity to the tumour before their planning procedure. Although the measurements of the fiducial markers in the images have a better contrast than those of soft tissues, it is still challenging to automatically localize the markers

[3]Compare, for example, 8×1.85 mm [129] cylindrical electromagnetic transponder to 3×0.8 mm [152, 174] radio-opaque marker, where the dimensions reflect "length × diameter".

due to *overlaying anatomical structures* appearing in the projection images and the *small size* of the markers[4] relative to the image field of view. For example, the images used in this thesis show measurements of the cylindrical markers of 3 mm long and 0.8 mm in diameter within a field of view that is approximately 27×27 cm^2. As such, a marker occupies around 0.03% of the image area. As an illustration, try to locate the markers in a part of a fluoroscopic frame shown in Figure 1.3 (*right*), which appears on page 6. It is quite challenging to do without specific training.

In this thesis, this problem is overcome by first localizing the fiducial markers in the 3D space by *segmenting* the markers in a better-contrast 3D CBCT image (Chapter 3) that was acquired during the same treatment session as the given fluoroscopic image sequence. The acquisition of the CBCT images at each treatment fraction is done according to a clinical target position verification protocol explained in Section 2.2. Then, the positions of the markers in the 2D space (the plane in which the 2D motion tracking is later performed) are estimated based on the known geometric relationship between the coordinate systems of the 3D CBCT and 2D fluoroscopic images. To account for intra-fraction motion that likely happened between the acquisition of the CBCT and the first fluoroscopic frame, as well as for possible inaccuracies of the system's geometric calibration, the images are aligned by applying *2D-3D image registration* (Chapter 5). With an optimal geometric transformation, found as a result of the registration, the positions of the markers are finally localized in the first fluoroscopic frame. After this initialization, the system proceeds to marker tracking by performing *2D-2D image registration* between the template formed from the first image of the sequence and the rest of the images (Chapter 6). In order to successfully perform both 2D-3D and 2D-2D registration procedures, the images are preprocessed by adapted and newly developed digital filters (Chapter 4). Having provided this broad outline of the marker localization and tracking procedure proposed in this thesis, we will now describe our contributions in more detail.

A typical approach to the registration of a 3D CT/CBCT image to a 2D x-ray projection/fluoroscopic image is to compute a *digitally reconstructed radiograph* (DRR), a 2D synthetic projection from the 3D image that mimics the geometric setup and intensity properties of the 2D image, and register the DRR to the 2D image, e.g., [42, 154, 249]. Often, the DRR generation involves significant computational resources. In addition, it can be challenging to register low-contrast images, in which important features, such as the marker measurements, are overlayed by the measurements of anatomical structures. To address these issues, the 2D-3D image registration approach proposed in this thesis generates the 2D image from the markers segmented

[4]For simplicity, we will use the term *marker* to refer to the *marker measurement*, or the *marker feature*, i.e., the graphical portrayal of the fiducial marker in the image. The term *fiducial marker* will be used to denote the physical marker object itself.

in the 3D CBCT image only, ignoring all other features found in the image. To distinguish this imaging modality, we introduce a new term – a *selective DRR*. Similar approaches of generating the 2D images only from *voxels* (volumetric pixels) whose intensity is above a certain threshold were presented in [7, 28]. In this thesis, the idea is extended by a number of procedures and criteria to determine whether the selected voxels actually belong to the markers or to other high-intensity features (explained in Section 3.1). Those belonging to other features are dismissed before the DRR generation.

In parallel with the construction of the selective DRR, the fluoroscopic images are preprocessed by digital filters, such as *temporal filtering* detailed in Section 4.1, and a newly designed *marker enhancement filter* proposed in Section 4.4. The application of the filters creates an image that contains measurements and possesses intensity characteristics similar to those in the selective DRR, making the process of registration efficient and reliable.

The proposed marker tracking algorithm is based on similar principles. In particular, the 2D-2D image registration between the first image and all other images of the sequence is used to quantify the motion. Instead of using the first image directly, we create a synthetic image from the markers that have been found by the localization procedure, which is analogous to the selective DRR computation. At the same time, all other images of the sequence are preprocessed by temporal filtering, followed by a simplified version of the marker enhancement filter. In addition to improving image quality by denoising and amplifying the markers, this approach incurs a much lower computational cost than the conventional method of registering discrete images that involves interpolation, as explained in detail in Section 6.1.

The key contribution of this thesis is the construction of the complete automated marker localization and tracking system. In the process of its development, a number of novel techniques have been proposed, including:

- **Method for computation of the 3D marker models.** As the CBCT images are computed from a discrete set of x-ray projections acquired during the rotation of the LINAC around the patient, they are often prone to reconstruction artifacts, in particular, *metal streak artifacts* [97, 151, 18, 248], around high-intensity markers. Artifacts can be described as sets of voxels in a shape of streaks possessing brightness comparable to the markers but, unlike markers, they do not represent a measurement of any real object or material with high x-ray attenuation properties. These voxels, naturally, have the potential to propagate into the synthetically created projection images, such as conventional or selective DRRs. This situation is not desirable as it creates the potential for a reduction in registration accuracy. As it may be challenging to separate the

artifacts from the marker voxels purely based on their intensity values, it is proposed to exploit knowledge about actual fiducial marker shape and dimensions, and create *marker models*. The positions and orientations of the models in the 3D space are computed based on the segmented marker voxels as described in Section 3.4. The models are subsequently used instead of the segmented markers in one of the proposed selective DRR computation methods.

- **Novel image generation approaches.** The selective DRR is generated from either the marker voxels or the marker models. The use of the marker models is justified in the cases of average to high resolution CBCTs, where marker orientations in 3D can be resolved reliably, and the use of the models further improves the appearance of the markers in the generated 2D image. However, if the CBCT voxel size is large, or there are other considerations involved, such as running time, the marker models should not be used. The details are provided in Section 3.5. To respond to these different possibilities, several image generation methods have been proposed in this thesis:
  - *Marker model DRR.* This method simulates an x-ray image creation process, where a pixel intensity in a 2D projection image depends on attenuation properties of the material that is placed between the corresponding elements of the detector and an x-ray source. Mathematically, this image generation method is based on a concept of *raytracing* [120]. In general, raytracing is very expensive computationally, and can only be used if the number of points is limited, such as the case with the marker models. The details are provided in Section 5.1.1.
  - *Voxel splat DRR.* If the marker models cannot be used but it is still desirable to include available orientation information into the generated image, for example, to study the rigidity of the organ over time, all marker voxels can be used to create a voxel splat DRR as explained in Section 5.1.2. Although the image is generated from separate voxels, this method creates contiguous features from contiguous sets of voxels, such as the segmented marker. This property becomes especially important if the CBCT voxel size is larger than the size of the pixels in the fluoroscopic images. Also, this method can be applied if the fiducial markers are of irregular shapes, e.g., long flexible coils [7].
  - *Marker splat DRR.* In this case, no orientation information is used. The markers in the generated image are represented by round spots as explained in Section 5.1.3. This method is appropriate for low resolution CBCTs, and is very computationally efficient.
  - *2D marker splat image.* This image generation method was designed for

marker tracking. It is the 2D analogue of the marker splat DRR defined above. Its use results in low computational cost of 2D-2D image registration as discussed in Section 6.1.

- **Novel marker enhancement filter (MEF) and its variations.** The MEF is a non-linear digital filter specifically designed for enhancement of the markers and suppression of the noise and anatomical features[5] in fluoroscopic images. The MEF amplifies features of certain sizes (scales) by analyzing the local structure of the image. Although the principles of local intensity analysis and linear scale space have been explored before, in the field of medical image processing they have been applied mostly in the enhancement of long line-like features, such as vessels, nerves, and bronchi [132, 188, 63]. In this thesis, the idea is adapted to the case of the markers, which are quite short, being only 3–4 times as long as their diameter. In brief, the MEF is built upon two images, where the first one amplifies the positions where the markers can be found with high probability, while the second one amplifies marker shapes found within these positions. The MEF also includes a novel *contrast enhancement function*. The details of the MEF design are provided in Section 4.4.

  Due to availability of additional information from the localization procedure, a simplified version of the MEF is used in tracking, which allows to achieve high computational efficiency as described in Section 6.1.

In addition to the necessity of the marker enhancement, which is very important for fluoroscopic images as they contain multiple non-marker features that can mislead the registration, the application of the MEF or its simplified version and the image generation approaches listed above has other important advantages. Although initially the images belong to different modalities (CBCT and fluoroscopy), and different input data is used in registration (marker positions in 3D or 2D versus a fluoroscopic image), the proposed MEF preprocessing and image generation approaches enable us to create images that have similar properties and appear to have almost the same modality. Indeed, their intensity characteristics and objects displayed in the images (markers) are quite similar. This creates the potential for successful use of the modelling and computational techniques that are typical for *single-modality registration*. *Multi-modality registration*, on the other hand, often requires the use of more flexible, sophisticated, and computationally expensive tools, which may require an involved interpretation of results. The use of the MEF and custom image generation also allows us to create images with minimal noise, and the registration of the smoother images is typically significantly faster and more reliable than of their noisy

[5] The term *anatomical features* refers to the measurements of anatomical structures in the images.

counterparts.

In addition to the novel modelling techniques described above, this thesis also presents a number of other innovations and thorough analysis of existing techniques with the purpose of designing, selecting, and adapting appropriate models and their parameters to incorporate them into our marker localization and tracking system. For example, a *model-based threshold* for the segmentation of the markers in the CBCT image was determined. It does not rely on the absolute intensity values of the voxels and, therefore, works well for different CBCT images. Next, the use of *hierarchical clustering* technique is an example of a justified selection of the system's component. It is utilized to divide the potential marker voxels into separate sets, each corresponding to a measurement of one entity (a marker or other feature), a necessary and important step before selective DRR generation. Both techniques are described in Section 3.1. Then, a number of heuristical procedures determining whether the clusters can be *classified as the marker or non-marker features* are presented in Sections 3.1 and 3.3. Further, the choice of the *temporal filter* for the fluoroscopic images and its parameters, discussed in Sections 4.1 and 5.4, arose from thorough testing with respect to the fluoroscopic image quality and temporal resolution of the filtered image sequence. The components of the image registration problem, such as the *distance measure*, *geometric transformation*, and *optimization technique* were selected among the pool of available tools as described in Sections 5.2, 5.3.2, and 5.3.3. The *bounds* on optimization parameters are determined based on a detailed analysis of the clinical literature presented in Section 5.3.1, where a large number of studies reporting on patient organ motion is summarized. The techniques to compute a reasonable optimization *starting point* that greatly increases the reliability of registration are presented in Section 5.3.3. After thorough testing of available approaches, we selected one of the known methods, and upgraded it to fulfil our task. Finally, the *prediction* part of a tracking model and its parameters were selected by analyzing the results presented in the literature and applying different models to our testing data sets as presented in Section 6.2.

In order to estimate the accuracy of mathematical models in their description of real-life processes or entities, it is important to perform *validation* procedures. Chapter 7 presents a detailed description of testing procedures with definitions of errors. The algorithms proposed in this thesis have been implemented as a Matlab software package (Matlab, The Mathworks, Inc., Natick, MA) called Gryphon, and tested on the CBCT and fluoroscopic images of physical phantoms and prostate patients. Validation results demonstrate that Gryphon can achieve submillimeter and subpixel accuracy in marker localization and tracking, which makes the techniques developed in this thesis suitable for clinical implementation.

This thesis addresses the full gamut of the conceptualization – design – implementation process starting from formulation of requirements, through development of model assumptions, derivation of mathematical models, discussion of algorithmic and implementation issues, and finishing with validation on phantom images and testing on patient data. As such, the writing style of this thesis may be viewed as an interleaving of scientific and engineering modes: Mathematical models that aim to hold some degree of generality and explore their options are followed by particular design choices pertaining to the application of fluoroscopic tracking to prostate cancer patients.

This thesis contains multiple forward references. First, the introductory parts of the chapters and sections often discuss methods at a high level, and include forward references to the details discussed later. We believe that such writing style is suitable to the broad range of backgrounds that the expected audience of this thesis will have. Professionals of different fields may find different selective parts interesting, and it was felt that a layered structure with appropriate referencing would be preferred over a linear storyline. Second, forward references are given to the experiments for the selection of parameters and validation of methods that are collected in Chapter 7, and implementation details that are provided in Appendix D, so as not to interfere with the main discussion in Chapters 3 – 6.

1.3 Thesis Outline

This introductory chapter is followed by the discussion of the particulars pertaining to motion tracking based on the fiducial markers, and explanation of intuition behind suggested filtering, localization and tracking methodologies in Chapter 2, which also establishes the mathematical foundation for the following discussion. Chapters 3 – 6 provide modelling and algorithmic details on each step of the 3D segmentation, filtering, 2D localization, and tracking procedures, respectively. Next, Chapter 7 presents the accuracy validation results of applying the proposed algorithms to the testing image data sets of physical phantoms and patients. An example of the analysis that can be performed on patient images in clinical setting using the methods proposed in this thesis and implemented in the Gryphon software is given in Chapter 8. Conclusions and possible extensions are presented in Chapter 9.

The thesis contains a number of appendices. First, software and hardware details are given in Appendix A. Then, though the details of the geometric transformation functions used in this thesis can be deduced, in a straightforward way, from the description of the coordinate systems in Section 2.4.2, explicit statements are provided in Appendix B for the convenience of the reader, and selected detailed

formulations are given in Appendix C. Finally, implementation details and algorithm listings are presented in Appendix D.

Chapter 2

Problem and Solution Overview and Preliminaries

This chapter begins with a discussion on the use of the fiducial markers in radiation treatment process and their types, and a literature review on marker based organ motion analysis in Section 2.1. This is followed by a description of the problem setup in Section 2.2, which includes a high-level overview of the architecture of the proposed system that consists of marker localization, image filtering, and marker tracking modules implemented in the Gryphon software. Also, the modelling assumptions are discussed. Following that, Section 2.3 briefly describes data sets with images used for testing of the algorithms proposed in this thesis. The chapter is concluded with a formal introduction to the modelling that will be employed, such as image model formulation, and the description of the geometric systems associated with the images used in this thesis in Section 2.4.

2.1 Organ Motion Analysis Based on Fiducial Markers

2.1.1 Use of Fiducial Markers for Target Localization

In conventional clinical treatment, an initial (external) setup of the patient on a medical LINAC is performed by the use of immobilization devices and by aligning patient's skin marks to the in-room lasers defining the LINAC's coordinate system. At Princess Margaret Hospital, this is followed by a CBCT image acquisition to enable visualization of the patient's inner structures. This daily CBCT is then compared to the planning CT image by the means of image registration to verify the target position, and the patient's position is adjusted if necessary. *Insufficient soft-tissue contrast* in the images can complicate image registration and, thus, can be an obstacle for a

patient daily setup. To mitigate this effect, metallic *fiducial markers* are often inserted into, or in close proximity to, the target. The markers produce measurements of higher contrast in the images compared to soft tissues due to the high x-ray attenuation properties of the material, and their positions become surrogates for the target position throughout the whole planning and treatment process.

In the proposed automated localization and tracking procedures, we will take advantage of the markers. As was explained in Section 1.1.5, assessment of internal organ motion through the fluoroscopic tracking of radio-opaque fiducial markers is a crucial component of various tasks associated with reduction of the uncertainty of the target position during radiation treatment. Fiducial markers appear as bright features in the attenuation x-ray and CT/CBCT images and can be in the shape of *cylinders*, *spheres*, long flexible *coils*, or spiral *stents*. The latter two types, however, are rarely used. Some examples of the marker dimensions are 3×0.8 mm [152, 174], 3×1 mm [98, 43], and 5×1.2 mm [141] for cylinders (length×diameter), 1.57 to 4 mm [141] or 2 mm [198] for spheres (diameter), $3-4$ cm×0.35 mm for coils [7], and 7 mm spring in diameter with the maximal expansion of the collar to 14 mm for stents [39].

All phantom and patient images used in this thesis display the measurements of gold cylindrical fiducial markers of 3×0.8 mm, as demonstrated in Figure 2.1. These markers are used clinically for daily target position verification, in particular, for prostate patients. Although models for the spherical markers are provided as well, to demonstrate the generality of the proposed approaches, most of the parameters, as well as validation and tracking results, are only presented for the aforementioned cylindrical markers.

2.1.2 X-Ray Tracking Systems in Literature

The literature on target localization and tracking performed by analyzing kV and MV x-ray images and CBCT projections is quite broad, and ranges from the description of separate parts and particular techniques to complete tracking systems. To place our work into the context, in this section we provide a brief overview of the existing systems, and the literature pertaining to specific methods is deferred to appropriate chapters and sections.

Several tracking systems combining hardware for image acquisition and software for image analysis, that perform tracking based on radio-opaque fiducial markers, have been reported in the literature:

- The Mitsubishi/Hokkaido fluoroscopic real-time tumour tracking radiotherapy system (RTRT) presented by Shirato *et al.* [198] uses four x-ray imaging systems

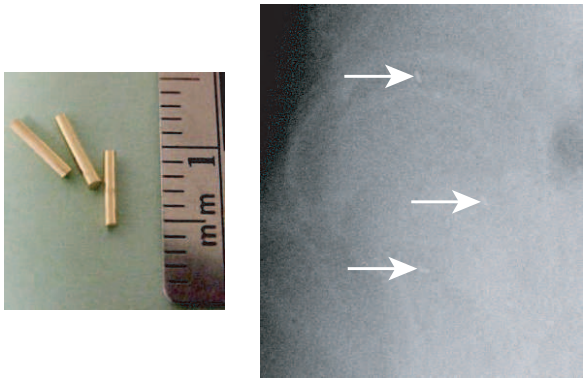


Figure 2.1: Gold fiducial markers (*left*) and their measurements in a fluoroscopic image of a prostate patient pointed by white arrows (*right*). As can be seen, the markers possess low contrast with the background in fluoroscopic images, making the automatic, and even manual, localization challenging.

fixed in the room around the LINAC. The system tracks one 2 mm gold spherical marker in real time, i.e., the images are processed on the fly, and the marker 3D position is reported every 0.033 seconds, with the accuracy of ± 1 mm as measured in the phantom experiment. Being synchronized with the LINAC, the system is used for gating, i.e., the irradiation is only triggered when the target is located in a predefined range of acceptable treatment positions.

- In the Accuray Cyberknife robotic radiosurgery system described by Schweikard *et al.* [193], the treatment beam is moved by a robotic arm to follow a moving target. To reduce the imaging dose to the patient, instead of using real-time x-ray imaging, the system tracks the 3D positions of four gold spherical fiducial markers 2.5 mm in diameter by combining occasional x-ray images with real-time tracking of the infrared emitters placed on the patient's chest or abdomen. The correlation between the motion of the infrared markers and the 3D internal marker positions, computed based on images from two x-ray imaging systems, has to be established prior to treatment. The model is continuously updated during the treatment. The described system does not use prediction, and experimental results show that for typical respiratory velocity of 2–4 mm/sec the beam lag is between 1 and 2 mm.
- A number of combined kV/MV imaging systems [141, 43, 204] compute the 3D positions of the fiducial markers from 2D positions found in MV and kV x-ray images. The kV x-ray imaging system is mounted so that the MV and kV beams are orthogonal to each other. Mao *et al.* [141] used stainless steel ball bearings from 1.57 mm to 4 mm and gold cylinders of 5×1.2 mm in size. The reported

positioning errors are within 1 mm. Cho *et al.* [43] synchronized the kV/MV imaging system with the dynamic MLC for tumour tracking. The time delay of the proposed tracking system is approximately 450 ms, and a tracking error for one 3×1 mm gold marker with the use of prediction is 0.9 ± 0.5 mm for the phantom experiment. In application to prostate patients, Slagmolen *et al.* [204] reported the 3D positioning accuracy for four 5×1 mm markers to be 0.5 ± 0.5 mm.

- Berbeco *et al.* [23] experimented with the number of kV imaging systems mounted on the LINAC. They observed that for one system mounted so that the MV and kV beams are orthogonal to each other, positioning errors in the beam's-eye-view for seven lung cancer patients, averaged over 36 evenly spaced gantry angles, ranged from 0.9 ± 0.3 to 7.5 ± 3.0 mm and the spread in the 99% range was 2.2–42.7 mm. The marker positions in 3D were derived from their 2D positions observed in the images by using the previously collected information about the 3D trajectory and assuming the repeatability of the relationship between the 2D and 3D positions. Berbeco *et al.* concluded that one imaging system was not adequate to resolve 3D marker positions with sufficient accuracy. Therefore, they proposed to use two kV x-ray imaging systems for the integrated radiotherapy imaging system (IRIS) created by the Massachusetts General Hospital and Varian Medical Systems, Inc. (Palo Alto, CA). The imaging units were mounted symmetrically on either side of the LINAC's head with 90° between the kV beams.
- Adamson and Wu [7] used one CBCT imaging system mounted on an Elekta Synergy LINAC to measure the 3D motion of three gold coils for prostate patients undergoing IMRT treatment. The 3D motion was computed by combining 2D marker positions found at several gantry angles. The combined 2D marker localization accuracy for three patients was 0.06 ± 0.35 mm and 0.04 ± 0.71 mm in the two coordinate directions. In 3D, the accuracy was -0.04 ± 0.3, 0.09 ± 0.36, and 0.03 ± 0.68 mm in the three coordinate directions. It was found that while drifting motion can be generally observed using the proposed tracking method, the oscillating motion, except for the very low frequencies, can not be accurately estimated.
- Ng *et al.* [164] used a LINAC with one x-ray imaging system mounted so that the MV and kV beams are orthogonal to each other to design a method for 3D tracking of three 3×1 mm fiducial markers in prostate patients. The kilovoltage intrafraction monitoring (KIM) approach uses a previously acquired set of CBCT projections to build 3D Gaussian probability density functions of the marker positions, that are later used to resolve the out-of-plane marker motion.

The in-plane positions are obtained from the kV x-ray images acquired during the treatment. The mean 3D marker positioning accuracy of KIM for ten patients ranged from 0.25 mm to 0.86 mm, with an average for all patients at 0.46 ± 0.58 mm.

The project described in this thesis originated from the desire to quantify intra-fraction motion of a patient or a cohort of patients. While it may be useful to track and quantify the motion in 3D, for many biological sites the PTV is designed by isotropically expanding the CTV borders, as was explained in Section 1.1.4. Therefore, among all possible directions in 3D, it may be the most desirable to quantify the *largest motion*. As will be discussed in Section 2.2, the prostate's smallest motion is typically observed towards the patient's side. Consequently, the prostate's greatest motion can be visible on the 2D x-ray images acquired from the patient's side (lateral images). Hence, we decided to use an already available Elekta Synergy LINAC with a *single kV x-ray imaging unit* mounted so that the MV and kV beams are orthogonal to each other (Elekta Oncology Systems, Atlanta, GA; This system is similar to another commercially available LINAC with one imaging unit, on-board imager (OBI) developed by Varian Medical Systems, Inc., Palo Alto, CA). Such systems are currently used clinically in the Princess Margaret Hospital to produce CBCTs for daily patient treatment setup and are supplied with the XVI software for image acquisition [92, 152]. The system with one imaging unit is cheaper in comparison to the one with multiple units, and simpler, meaning it is easier for production, daily operation, and maintenance. Whether it will be possible to successfully resolve the out-of-plane motion with such configuration (for example, through analyzing the relative positioning of multiple markers) is a question to explore, but it should be possible to track the in-plane marker motion with sufficient accuracy. Another important difference between the majority of the available systems and our proposed configuration is the use of *smaller fiducial markers*. While smaller markers can offer reduced probability of complications related to marker insertion, their automatic localization and tracking is more challenging. Finally, the *lateral images* typically possess lower contrast and higher noise comparing to the x-ray projections obtained from any other angle due to the body anatomy. A number of tracking systems validated on patients or humanoid phantoms report the decrease in tracking accuracy or the increase in failure rates for the lateral images, e.g., [7, 15, 23].

Taking into account the differences in the technical configuration with the systems described above, such as the use of a single imaging unit, small markers, the necessity to track in the most challenging low contrast projections, and the desire to be able to batch-process multiple data, we have identified the following general directions for improvement and novelty:

- **Image enhancement.** As the projection x-ray images possess low contrast (see Figure 2.1), in part, due to the limited x-ray imaging dose, it is typically necessary to preprocess them before any marker detection can be reliably performed. Although there are many noise reduction filters available, there is a particular need for a *specific marker enhancement method*, that can efficiently and effectively amplify the markers and suppress other measurements in the images.
- **Automation and robustness.** In the proposed tracking system, we aim to limit the need for user interaction as much as possible by selecting models and parameters that either work for multiple data, or by creating mechanisms for automatic selection of parameters by the system itself. We perform extensive validation to not only quantify the accuracy of marker localization and tracking, but also to demonstrate that the system produces good results for ranges of parameter values, and for different data.
- **Computational efficiency and reliability.** Although computer hardware improves every year, computational efficiency of the algorithms remains a concern, especially when working with large data, such as sets of images. While the use of supercomputers, distributed systems, and graphic processing units (GPU) can alleviate the problem, we believe that it is best to take a resource-savvy approach. One side of it is to design efficient algorithms and implementation, but the other is to reduce the volume of data that needs to be processed. By carefully using the data collected prior to tracking, such as other images of the patient and geometric calibration parameters, we are able to select very specific regions of the images. In addition to reducing computational time, this approach also improves reliability of the system by increasing the probability of obtaining the expected results and minimizing the chance of failure.

The following chapters address specific methods designed with these ideas of novelty and improvement in mind, and the general scheme of the localization and tracking approach is presented in the next section.

2.2 Problem Setup, Assumptions, and Overview of Approach

We aim to characterize the *intra-fraction motion* of a target during some period of time by analyzing 2D fluoroscopic image sequences.[1] As explained in Section 1.1.3,

[1]Sometimes, the sequences of 2D images can be regarded as the 3D data, with the time as the third dimension. The notation of the dimension that we use in this thesis always refers to the dimension of the image in space, regardless of whether the time scale is involved or not.

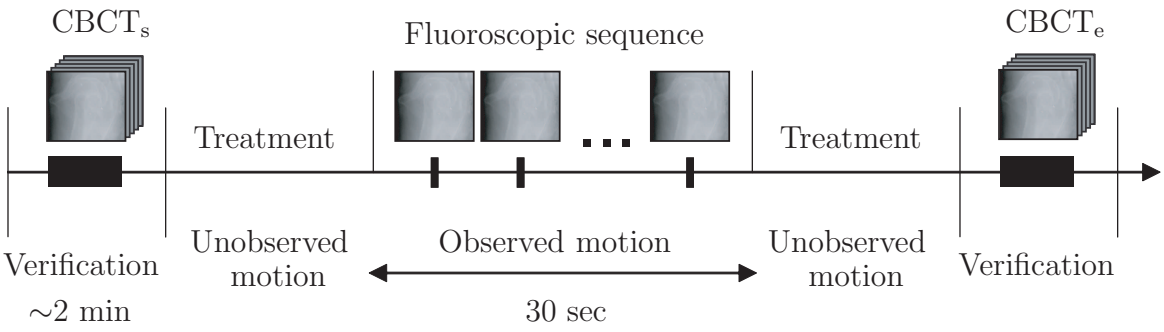


Figure 2.2: A protocol for a treatment session with the intra-fraction motion evaluation experiment.

a *fluoroscopic sequence* is a temporal series of 2D x-ray projection images acquired with a stationary kV imaging system. All image sequences used in this thesis were acquired at 5.5 Hz (frames per second) at a technique of 120 kVp, 32 mA, 40 ms per frame for phantom images, and 120 kVp, 64 mA, 20 ms for prostate patients. Typically, the patient sequences consist of approximately 165 frames (30 seconds of acquisition), while the phantom sequences differ in length depending on specifics of each experiment. Marker tracking enables us to quantify the displacement of the target in time, which creates an opportunity to make adjustments to the treatment, in particular, select an appropriate PTV, whether for a single patient or a patient cohort.

Figure 2.2 depicts a typical treatment session for a prostate patient with the *intra-fraction motion measurement experiment* included[2]. When the patient comes in, a verification image, denoted CBCT$_s$, is acquired and compared against the planning CT. After target position verification and correction (usually, by moving the couch), and sometimes another CBCT acquisition, the treatment commences. At some point in time during the treatment, typically when the machine has to be moved from one angle to another, i.e., the MV beam is naturally turned off in accordance with the treatment design, which typically involves irradiating from several angles, fluoroscopic image sequence is acquired. Then, the treatment continues and, at the end, another image, CBCT$_e$, is taken for verification: If the target exhibited significant displacement during the treatment procedure, it probably did not receive the prescribed radiation dose, in which case some adjustments to the treatment plan may be required.

To be reusable and applicable to a large number of patients, a motion tracking

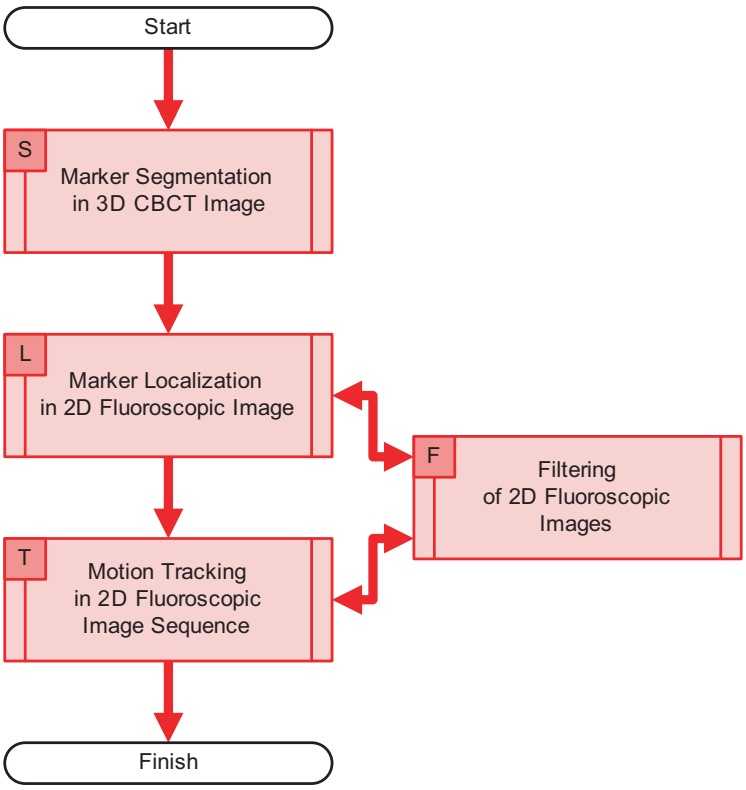
[2]That is, the motion measurement experiment is not a part of a standard treatment session, and was designed and incorporated into the sessions for a collection of target motion data in multiple patients.

method should be implemented as a software tool, which provides automated processing, and minimizes, or even eliminates, the need for user interaction. Although the markers are characterized with higher intensities than the measurements of the surrounding soft tissues, automatic marker detection is challenging due to overlaying anatomical features and noise in the x-ray projection images (See Figure 2.1). Often, the markers have higher intensity values only in comparison to their immediate neighbourhood, i.e., they possess *local* contrast in a small region rather than *global* in the whole image. Hence, a reliable marker detection procedure would require to conduct local intensity analysis, which can be expensive to perform on the whole image. Consequently, a definition of a smaller region of interest (ROI) within the image that contains the markers is desirable.

To overcome the aforementioned 2D detection issues, we propose to first find the markers in the CBCT image acquired during the same treatment session as the fluoroscopic sequence that is to be processed. The markers are typically easier to detect in the 3D CBCT images due to the absence of overlaying features and better contrast with the background. Then, knowing a geometric relationship between the coordinate systems of the CBCT and fluoroscopic images, approximate marker locations in 2D can be estimated. To account for the intra-fraction motion that likely happened between the acquisition of the CBCT and the fluoroscopic sequence (see Figure 2.2), as well as for a possible geometric miscallibration, the estimated positions are corrected using the 2D-3D image registration. This is only performed for the first image of the sequence to *initialize tracking*. Then, the tracking is performed by 2D-2D image registration between the first image with the localized markers and the remaining images of the sequence.

Flowchart 2.1 presents an overview of the proposed approach.[3] The boxes in the figure correspond to the algorithmic blocks, each of which will be addressed in the following chapters. The processing flow starts by localizing the markers in the 3D CBCT image using the segmentation procedure in block $\langle\!\langle \mathbf{S} \rangle\!\rangle$ as described in Chapter 3. Either of the two CBCT images are preferred to the planning CT because they are acquired during the same treatment session as the fluoroscopic sequence. Therefore, there is no need to correct for *inter-fraction displacement* or *deformation*. *Real-time tracking* would require the use of the $\mathrm{CBCT_s}$ image and the parameters of the couch correction (if available) that was likely performed as a result of the initial target position verification. In the case of *post-processing*, such as described in this thesis, we prefer to use the $\mathrm{CBCT_e}$ image, similar to the approach suggested in [7], to omit the necessity to compensate for couch correction. It is expected that $\mathrm{CBCT_e}$ gives a closer estimate of the marker locations than $\mathrm{CBCT_s}$.

[3] The flowchart notation is explained in Notation and Symbols section on page xxxi.



Flowchart 2.1: Overview of the proposed approach for fiducial marker localization and tracking.

The segmentation is followed by the 2D marker localization in block $\langle\!\langle \mathbf{L} \rangle\!\rangle$ described in detail in Chapter 5. It is performed by the means of 2D-3D image registration between the *selective DRR*, formed from the 3D markers found in block $\langle\!\langle \mathbf{S} \rangle\!\rangle$, and the first image of the fluoroscopic sequence, preprocessed in block $\langle\!\langle \mathbf{F} \rangle\!\rangle$. As can be seen from Flowchart 2.1, the filtering block $\langle\!\langle \mathbf{F} \rangle\!\rangle$ is called from both localization and tracking procedures. After the temporal and marker enhancement filtering is applied to the corresponding fluoroscopic images, as described in Chapter 4, the control is returned to the calling procedure.

Although the filtering procedures are designed and tuned in a way that amplifies the markers and suppresses the background, and in many cases the markers can be identified in the filtered images by a simple intensity threshold, some anatomical features can occasionally give a filter response similar to that of the markers, which may lead to false positives in 2D marker detection. The localization of the markers in the 3D CBCT followed by applying 2D-3D registration to the filtered fluoroscopic image addresses this issue and improves detection reliability.

After the markers have been localized in the first image of the sequence, marker tracking commences. The tracking procedure is referred to as block $\langle\!\langle \mathbf{T} \rangle\!\rangle$ in Flowchart 2.1, and is described in detail in Chapter 6. It is performed by the means of 2D-2D image registration between a *tracking template*, which is created from the markers localized in the first image of the sequence, and all subsequent fluoroscopic images, preprocessed by the filters in block $\langle\!\langle \mathbf{F} \rangle\!\rangle$.

Although we expect that the tracking system presented in this thesis can be used for multiple anatomical sites, here we specialize it to perform automatic measurement of intra-fraction motion in prostate cancer patients. A number of studies have been conducted to assess magnitudes of marker migration and prostate deformations by measuring *inter-marker distances* (IMD) during the course of treatment [36, 114, 176, 190]. It was concluded that the migration of the fiducial markers within the prostate is *negligible* and the positions of the markers represent an accurate and reliable surrogate of the prostate location during the whole treatment course. Variations in individual marker locations were attributed to deformation of the prostate due to daily anatomical changes in surrounding organs or as a reaction to the treatment itself, e.g., prostate shrinkage. Therefore, we make an assumption that should the fiducial markers exhibit intra-fraction displacement relative to each other between the acquisition of the fluoroscopic sequence and the CBCT within the same session, it is deemed negligible. In other words, in registration of the CBCT and the fluoroscopic images, we treat the fiducial markers as one object. Therefore, we assume that markers as a whole are only subject to *rigid transformations* (rotations and translations) during the time between fluoroscopic and CBCT acquisitions.

A number of studies have indicated that the prostate motion in patient's left-right (LR) direction is smaller than in anterior-posterior (AP) and superior-inferior (SI) directions (a review is provided in Section 5.3.1). For patients that follow a certain dietary protocol intended to reduce intra-fraction organ motion, the prostate tends to rotate in an orderly manner around its apex, and typically moves as a rigid body in a slow and constant motion, consistent with physiologic patterns [68]. All fluoroscopic image sequences analyzed in this thesis were acquired in a *lateral* direction, i.e., from a patient's side. Although lateral images are harder to analyze due to a typically lower contrast (the human body is usually wider in LR direction comparing to AP), such configuration gives an opportunity to measure the largest expected intra-fraction prostate motion among all directions in 3D, as the AP and SI directions appear in-plane. Apart from this particular clinical interest, a reliable method for marker detection can potentially improve tracking performance in other applications, as the reduction in detection and tracking accuracy in the lateral images has been repeatedly noted, e.g., [7, 15, 23].

The next section describes the phantom and patient image data sets that were used for testing of the proposed approach.

2.3 Testing Data Sets

Although all necessary details on the design of physical phantoms and phantom testing data sets are provided in Chapter 7, and patient data sets are described in Chapter 8, this section presents a brief introduction on the aforementioned topics necessary for understanding the discussion in Chapters 3 – 6. The data sets used in this thesis include sets of testing images with pertaining information, in particular, coordinate systems associated with the images. First, our proposed algorithms are tested on the images of two verification phantoms. When the design is finalized, we also perform validation on a testing patient data set containing multiple measurements of three prostate cancer patients.

2.3.1 Radio-Surgery Verification Phantom

The radio-surgery verification phantom (The Phantom Laboratory, Salem, NY, further referred to as the *RSVP*) is a plastic pelvic hull filled with water shown in Figure 2.3 (*left*). The phantom's top part has an opening with a spherical plug that can be used for mounting custom inserts inside. Such design allows to create rotations of the insert, while shifts can be performed by moving the LINAC's couch where the phantom is placed for imaging. Inside the RSVP, there are two long structures simulating bones.

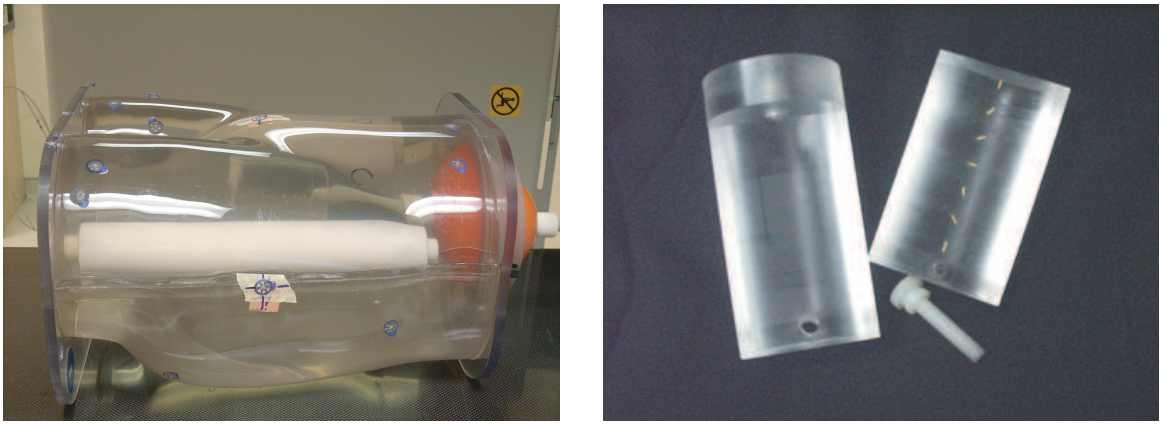


Figure 2.3: Radio-surgery verification phantom (RSVP, *left*) and a custom insert with seven gold cylindrical fiducial markers (*right*).

Figure 2.3 (*right*) shows an insert used for all RSVP measurements in this thesis. It consists of two half-cylinders that are fixed together with a screw. One half-cylinder has seven gold fiducial markers 3×0.8 mm. They are placed to have their axes in one plane, and their centres on one line, with 10 mm distance in between the centres of the neighbouring markers.

All RSVP measurements were performed with a stationary phantom aimed at the validation of the 3D and 2D marker localization, and the filtering procedures. The phantom was placed on a LINAC equipped with a kV x-ray imaging system, and imaged at different shifts and rotations. Altogether, nine fluoroscopic sequences (1024×1024 pixels) and nine corresponding CBCT projection sets (512×512 pixels) were acquired. The projections were reconstructed into the CBCT images with 0.5, 1, 2, and 4 mm voxels.[4]

2.3.2 Quasar Motion Phantom

The Quasar programmable respiratory motion phantom (Modus Medical Devices Inc., London, Ontario, further referred to as the *Quasar phantom*) shown in Figure 2.4 (*left*) consists of the following parts: a plastic piece in a shape of a patient's abdomen or chest, a set of inserts simulating different body tissues, and a Scottish yawk mechanism for moving the inserts. The phantom can produce sinusoidal motion (in time) along its vertical axis, where a period and an amplitude can be adjusted. In addition, the phantom can move the insert according to an arbitrary motion trace that is loaded

[4]When referring to cubic voxels, for simplicity we will always say, for example, 0.5 mm voxels, meaning the voxels of $0.5 \times 0.5 \times 0.5$ mm^3 dimensions.

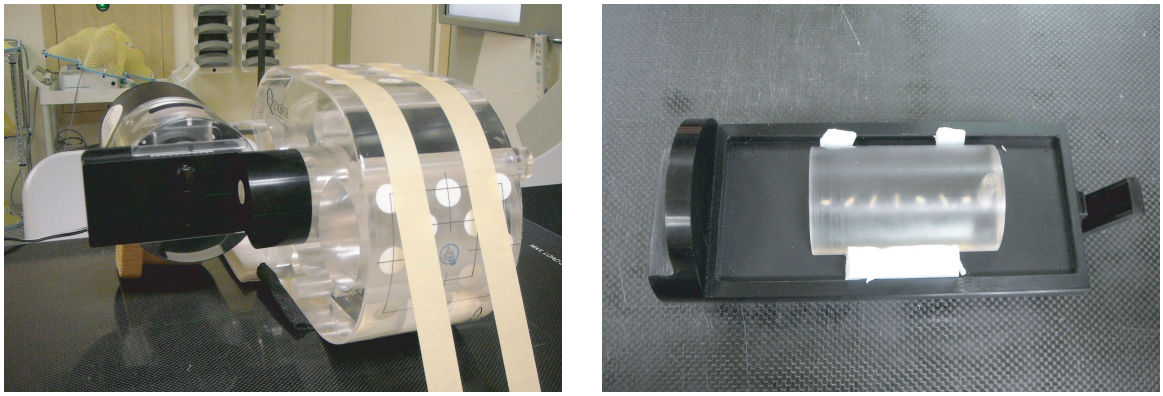


Figure 2.4: Quasar motion phantom (*left*) and an insert with seven gold cylindrical fiducial markers (*right*).

as a sequence of displacements.

Figure 2.4 (*right*) shows an insert that was used for the Quasar phantom in experiments described in this thesis. It combines half of a standard Quasar insert that attaches to the motion mechanism, and half of the plastic cylinder with the fiducial markers described in Section 2.3.1.

The Quasar phantom was used to acquire image sequences for testing of the filtering and marker tracking procedures. The sequences show cases in which the phantom is stationary, steadily moving at three different speeds, performs three traces simulating pseudo-regular breathing and irregular motion, and a sinusoidal motion with peak-to-peak amplitudes of 2, 4, 10, and 20 mm and periods of 2.5, 5, 10, and 15 seconds. Also, 5 CBCT images with 0.25 mm voxel size were reconstructed from projections of 1024×1024 pixels. These images were used to test the 3D localization procedure for high resolution, because it was not possible to reconstruct the 0.25 mm voxel CBCTs from the RSVP projections of 512×512 pixels. Although CBCT images with 0.5 and 1 mm voxels are often utilized clinically, there is an increasing demand to use 0.25 mm voxel images for some anatomical sites. It was possible to reconstruct the CBCT images of lower resolutions from the Quasar phantom as well. However, we chose to use the RSVP, as it was bigger in size comparing to the Quasar: As a result, its images had lower contrast, and better resembled the patient images.

2.3.3 Patient Testing Data Sets

Although testing of the algorithms and the assessments of their accuracies were performed on the phantom images, due to the possibility to create a controlled environment with the phantoms, we also used a patient testing data set to verify that the

procedures run smoothly, produce expected results, and that parameter values are reasonable. The major difference between patients and phantoms is the presence of multiple anatomical structures in the patients, which may result in the images with lower contrast and higher noise. On the other hand, patient images do not have parts such as phantom insert holders, screws, and surface fiducial markers.

The collection of patient data was approved by the University Health Network Research Ethics Board (UHN-REB) [3]. The measurements of Patient 1, 2, and 3 include 10, 10, and 5 fluoroscopic sequences, respectively, and daily CBCTs for the whole treatment duration, including the CBCTs produced during the fractions with the fluoroscopy acquisition. Patient 1 has 74 CBCTs with 1 mm voxels and 6 with 2 mm, Patient 2 has 75 images, and Patient 3 has 74 CBCTs, both with 1 mm voxels.

2.4 Modelling Preliminaries

This section presents mathematical formulations necessary for understanding the discussion in Chapters 3–6: an image model definition, and a description of the coordinate systems associated with the images and geometric transformations that will be used in this thesis.

2.4.1 Image Model

A *discrete image*, or image data, is a d-dimensional array of intensity values $\mathtt{A} = (\mathtt{A}_{i_1,\ldots,i_d}) = (\mathtt{A}_i)$, where $i_k = 1,\ldots,m_k$. The size of the discrete image is $m_1 \times \ldots \times m_d$, and, in general, $\mathtt{A}_i \in \mathbb{R}$. The discrete image represents a measurement of some properties of an object or a set of objects, such as spectra of reflected light for a conventional photography, or x-ray attenuation of the object's material in x-ray images. Medical images are usually 2D, 3D, or 4D (3D plus time). Due to the fact that the separation between the intra- and inter-fraction motion is important in a clinical setting, at times a data set can be referred to as 5D, meaning a set of 3D images that contains both intra- (4^{th} dimension, time within the fraction) and inter-fraction measurements (5^{th} dimension, time between the fractions).

The elements of $\mathtt{A}$ are often associated with a rectangular *grid* of points $\check{p}^i \in \Gamma^d \subset \mathbb{R}^d$. Such knowledge helps to relate the measurements to some coordinate system, and, as a result, to a measured object or a patient. Hence, each value $\mathtt{A}_i = \mathtt{A}[\check{p}^i]$ is an intensity corresponding to a d-dimensional rectangular *cell* i, called a *pixel* in 2D images and a *voxel* in 3D, which is centered in a grid point $\check{p}^i \in \Gamma^d$. If ξ_k is the sample spacing of grid Γ^d along the k-th dimension, the cell i is defined as a set of points $\{q \mid (\check{p}^i_k - \xi_k/2) \leq q_k \leq (\check{p}^i_k + \xi_k/2),\ k = 1,\ldots,d\}$. In x-ray projection images,

the value $\mathtt{A}[\check{p}^i]$ relates to the number of photons that reach a detector element, i.e., it can be treated as an integral value of all intensity values measured at points within the corresponding cell i. More information and examples of grids can be found in Modersitzki [150].

The grid can be *isotropic* or *anisotropic*, where isotropic pixels/voxels refer to equally sized pixels/voxels in all dimensions, i.e., squares and cubes, while anisotropic pixels/voxels can have different dimensions. The anisotropic grids are primarily used in 3D images with anisotropic voxels, when the image consists of a set of 2D slices with isotropic pixels but the distance between the slices is typically larger than the pixel size, e.g., [188]. Although usually *regular*, the sample spacing of the grid can sometimes be *irregular* in order to measure a particular region of interest in more details than the rest of the scene.

At our perception scale we consider a measured object to be a continuous entity. The fact that the measurement of properties of this continuous entity results in a discrete image leads to two conceptually different approaches of defining an *image model*: as a continuous intensity function or as a mapping defined for a grid of points. Both approaches are used in the literature, e.g., Modersitzki [149] gives a continuous definition and Hill *et al.* [83] a discrete one, and many authors implicitly switch between the discrete definition (a measured image data array) and the continuous model (interpolation) in situations when the intensity values between the grid points have to be estimated, e.g., [34]. Following Modersitzki [149, 150], we define an *image* as a continuous intensity function:

$$\mathcal{A} : \Omega \subset \mathbb{R}^d \to \mathbb{R}, \tag{2.4.1}$$

i.e., every point $p \in \Omega$ is assigned some intensity value $\mathcal{A}(p)$, where Ω is a bounded set. The following assumptions are accepted in this thesis: (i) all images are *greyscale*, or *monochromatic*; (ii) the image is not defined outside of Ω, which does not affect the models presented in this thesis as the images typically show a scene larger than the object of interest, and it is not required to operate outside Ω but rather on a region of interest within Ω.

A continuous image can be computed from image data $\mathtt{A}$ by *interpolation*: $\mathcal{A}(p) = \mathcal{I}(\mathtt{A}, \Gamma^d, p)$. The most common choices for $\mathcal{I}$ in the context of medical image registration include nearest neighbour, linear, quadratic, and cubic spline interpolation [69, 149]. Depending on the choice of $\mathcal{I}$, $\mathcal{A}(\check{p}^i)$ does not necessarily equal $\mathtt{A}[\check{p}^i]$, $\check{p}^i \in \Gamma^d$. Although at first it may be viewed as the potential for increasing the error, it can be justified by the assumption that the measured object is, in general, continuous and smooth, and the digital measurement is typically quantized, which can add measurement errors in addition to other physical processes that cause noise during image

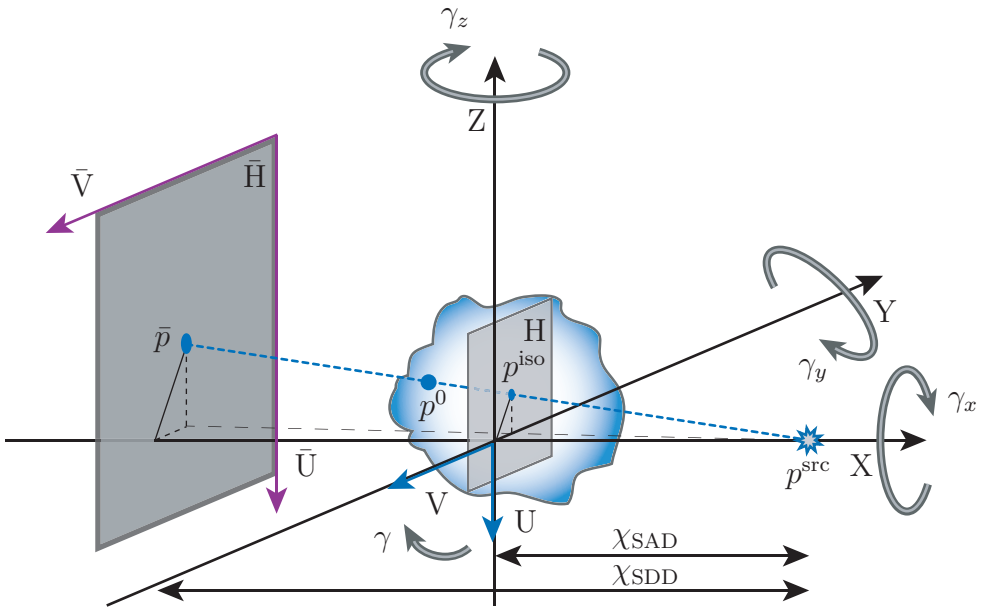


Figure 2.5: Coordinate systems associated with the patient and 3D images (XYZ), and the 2D images ($\bar{\mathrm{U}}\bar{\mathrm{V}}$ and UV).

acquisition. There is also a practical reason behind: Image enhancement and registration methods typically perform better on images possessing certain characteristics, such as smoothness.

In this thesis, we typically use continuous images when we describe mathematical models, and discrete for algorithms and implementation.

2.4.2 Coordinate Systems and Geometric Transformations

All images used in this thesis come with information about spatial location of their grid points in the corresponding coordinate systems. In this thesis, we use three coordinate systems called XYZ, UV, and $\bar{\mathrm{U}}\bar{\mathrm{V}}$ depicted in Figure 2.5:

- XYZ is a 3D right-handed coordinate system associated with the CBCT image and the patient. It is fixed to the room with the origin at the LINAC's centre of rotation. The coordinates of the grid points are assigned by the XVI software during volumetric image reconstruction from 2D x-ray projections. The units are millimeters.
- $\bar{\mathrm{U}}\bar{\mathrm{V}}$ is a 2D left-handed coordinate system of the flat-panel detector that is used for the x-ray projections and fluoroscopic images. The units of $\bar{\mathrm{U}}\bar{\mathrm{V}}$ are pixels, i.e., integer numbers either spanning 0 to $m_k - 1$ or 1 to m_k, where $m_1 \times m_2$ is the size of the discrete image. We use 1 to m_k numbering as this is the convention for the Matlab arrays. The position of plane $\bar{\mathrm{H}}$ on which $\bar{\mathrm{U}}\bar{\mathrm{V}}$ is defined is not

fixed in the XYZ system: It rotates with the flat-panel as the LINAC rotates around the patient.

- The images in the $\bar{\text{U}}\bar{\text{V}}$ system do not provide spatial information about the measurements, and the marker motion has to be quantified in millimeters, in relation to the geometry of the patient's organs. To achieve this, we introduce a third coordinate system, called UV. It is a 2D left-handed system with the origin at the LINAC's centre of rotation defined on plane H, where the planes H and $\bar{\text{H}}$ are parallel. We call H an *isocentric plane*. The system's units are millimeters. The positions of the points in UV are computed based on their positions in $\bar{\text{U}}\bar{\text{V}}$ and the knowledge of the image acquisition geometry including calibration information. Geometric transformations in XYZ, $\bar{\text{U}}\bar{\text{V}}$, and UV systems, as well as the transformation from $\bar{\text{U}}\bar{\text{V}}$ to UV, are detailed in Appendix B.

The point p^{src} in Figure 2.5 corresponds to the position of the imaging kV x-ray source, which can be treated as a point source. Matter present in the spatial location corresponding to $p^0 \in \mathbb{R}^3$ (XYZ) is penetrated by the diverging x-rays. The measurement of its x-ray attenuation, together with the measurements in all points along the line going through the points p^{src} and p^0, is recorded at point $\bar{p} \in \bar{\text{H}}$ on the flat-panel detector ($\bar{\text{U}}\bar{\text{V}}$). By performing transformation from $\bar{\text{U}}\bar{\text{V}}$ to UV, we find point p^{iso} (UV).

When the LINAC rotates around the patient, the point p^{src} and the planes $\bar{\text{H}}$ and H rotate around the Y axis. The rotation angle ϕ is recorded during image acquisition by the XVI software, and $\phi = 0°$ corresponds to the position for which the kV x-ray source is on the positive part of X axis, denoted by X+, as shown in Figure 2.5, and is increasing in the clockwise direction according to IEC gantry angle convention [5]. Axes $\bar{\text{V}}$ and V are parallel to Y and have a direction opposite to Y, while $\bar{\text{U}}$ and U rotate together with H and $\bar{\text{H}}$, and align with the negative Z direction when $\phi = 0°$. The distances from p^{src} to H (χ_{SAD}, source to axis distance) and to $\bar{\text{H}}$ (χ_{SDD}, source to detector distance) are known from the technical setup, and the values used in our experiments are given in Table B.1.

In our geometrical setup, we assume perfect alignment. However, any physical system comes with certain *geometric inaccuracy*. For example, the flat-panel detector can be slightly tilted, and values of calibration parameters and rotation angles are subject to the measurement and recording accuracy of the hardware and software. These misalignments can to some extent be measured with specially designed phantoms, however, it can not always be performed in a real-life situation. We will provide more discussion about this issue later in this thesis and explain how the design of our model is immune to slight geometric miscalibration in Section 5.3.

As explained in Section 2.2, we assume that the markers are only subject to

rigid transformations. The detailed definitions of geometric transformations are given in Appendix B, and here we only provide short explanation and a list of notations that will be used in this thesis. Note that before geometric transformations are performed, the points are expressed in *homogeneous coordinates* [238], i.e., the points are expressed as the vectors with one extra coordinate, e.g., $p = (p_x, p_y, p_z)^T$ becomes $\tilde{p} = (p_x, p_y, p_z, p_w)^T$, where $p_w = 1$. After performing necessary transformations, the points are transformed back to their initial dimensions by dividing by the last coordinate, and discarding it, e.g., $\tilde{q} = (q_x, q_y, q_z, q_w)^T$ becomes $q = (q_x/q_w, q_y/q_w, q_z/q_w)^T$, $q_w \neq 0$. The use of homogeneous coordinates allows to have a unified representation for rotation, translation, and perspective transformation as 4×4 matrices for the 3D, and as 3×3 matrices for the 2D coordinates. In the discussion below, we will use the tilde to denote the points in homogeneous coordinates.

The notation used for geometric transformations is the following:

- $\mathcal{T}^{3D}$ – 3D rigid transformation in the XYZ system. A transformed point $q \in \mathbb{R}^3$ is obtained from $p \in \mathbb{R}^3$ as:

$$q = \mathcal{T}^{3D}(a, p), \text{ or } \tilde{q} = R^{3D}T^{3D}\tilde{p}, \tag{2.4.2}$$

  where $\mathcal{T}^{3D} : \mathbb{R}^6 \times \mathbb{R}^3 \to \mathbb{R}^3$ is the transformation with six parameters $a = (t_x, t_y, t_z, \gamma_x, \gamma_y, \gamma_z)^T$ (three shifts and three rotation angles), and R^{3D} and T^{3D} are the 4×4 matrices of 3D rotation and translation, respectively, defined in Appendix B.1.
- $\mathcal{T}^{2D}$ – 2D rigid transformation in the $\bar{U}\bar{V}$ and UV systems. An application of $\mathcal{T}^{2D} : \mathbb{R}^3 \times \mathbb{R}^2 \to \mathbb{R}^2$ to a point $p \in \mathbb{R}^2$ results in:

$$q = \mathcal{T}^{2D}(b, p), \text{ or } \tilde{q} = R^{2D}T^{2D}\tilde{p}, \tag{2.4.3}$$

  where $b = (t_u, t_v, \gamma)^T$ is a vector of parameters (two shifts and one rotation angle), and 3×3 matrices of 2D rotation and translation R^{2D} and T^{2D} are defined in Appendix B.2.
- Some models require the use of inverse rotation and translation transformations, R^{-1} and T^{-1}, the definitions of which can be found in Appendix B.3.
- The transformation from $\bar{U}\bar{V}$ to UV is only performed once in the beginning for all fluoroscopic images, and all models and algorithms presented in this thesis assume that the points are defined in the UV system. Appendix B.4 contains the corresponding transformation equations.
- In the process of 2D-3D image registration, it is required to project the 3D points whose coordinates are expressed in the XYZ system (CBCT image) to the plane H, and express their coordinates in the UV system (fluoroscopic image with the grid transformed to UV). The perspective transformation from XYZ to UV is denoted by $\mathcal{P} : \mathbb{R}^3 \to \mathbb{R}^2$, and is derived in Appendix B.5.

Part I

Mathematical Modelling and Algorithm Design

Chapter 3

Marker Segmentation and Modelling in 3D Space

This chapter describes the segmentation and modelling of the markers in 3D space, that is the XYZ coordinate system associated with the CBCT image. This is the first step in the proposed localization and tracking method, denoted by the label $\langle\!\langle \mathbf{S} \rangle\!\rangle$ in Flowchart 2.1. The methods described here can also be used separately for marker-based analysis of the CT/CBCT images. Examples include computation of inter-marker distances (IMD) to detect organ deformations over time, and inter-fraction marker displacements to verify the accuracy and reproducibility of the patient treatment setup. Applications of both of these cases to a testing patient data set (described in Section 2.3.3) are presented in Chapter 8.

Approaches to using 3D images to either predict the positions of the markers in the 2D images, or to select a search ROI in the 2D images, were previously reported in the literature [7, 60, 141, 169, 204]. To measure marker motion in fluoroscopic sequences of the prostate patients, Adamson and Wu [7] performed image registration between fluoroscopic images and a 2D image constructed from the markers segmented in the CBCT image. The ROI in the CBCT was selected manually, then the markers were detected by using an intensity threshold. Fledelius *et al.* [60] used the 2D projection images, from which the pretreatment CBCT was reconstructed, to create 3D maker models consisting of size, orientation, and estimated 3D trajectory during the CBCT scan. Then, they projected the models to the 2D space to assist with 2D marker segmentation in real time. Mao *et al.* [141] and Slagmolen *et al.* [204] formed a search ROI in the 2D images as a region around the predicted 2D marker positions that were found by projecting the markers segmented in the planning CT by an intensity threshold. Similarly, Park *et al.* [169] selected a ROI in a 2D image by using a DRR of the planning CT.

As can be seen from multiple studies, marker localization or segmentation in

3D is typically performed by applying an *intensity threshold* to the CT [204, 141, 161] or CBCT images [151, 152, 7]. The result of universal, or global, thresholding is a binary image, where 1's are assigned to the points with the intensity larger than, or larger or equal to, some threshold value θ, and 0's otherwise[1] [180]. Alternatively, it is possible to set to 0 only the intensity values below θ but keep the original values above θ unchanged.

While thresholding with a predefined value θ may work well for the CT, where the CT intensity values relate to the x-ray attenuation of the material, and thus, the measurements of different materials tend to be represented by different *known* CT numbers, applying the same strategy to the CBCT images may be challenging, as the images can have different *scaling*. For example, Figure 3.1 demonstrates intensity distributions of the points in the volume of interest[2] (VOI) containing the markers. The results are shown for four CBCT images with the voxel sizes of 0.5, 1, 2, and 4 mm reconstructed from *the same* set of the CBCT projections displaying the RSVP, where the description of the phantom is given in Section 2.3.1. Although the most of the intensity values fall into the range between 750 to 1,150 intensity units, the overall ranges of the VOIs are quite different: $-1,646$ to 14,383 for the image with 0.5 mm voxels, 320 to 5,660 for 1 mm voxels, 445 to 2,088 for 2 mm voxels, and 733 to 1,407 for 4 mm voxels. While the selection of pixels with the intensity of at least 6,000 results in successful marker segmentation in the CBCT with 0.5 mm voxels, it is not even valid for the rest of the images, as 6,000 falls outside of their ranges.

In addition, it is possible that measurements of other objects in the VOI possess CBCT intensities similar to those of the markers. Based on our empirical observations and general understanding of the CT/CBCT image formation, the occurrence of non-marker measurements that have similar intensities to those of the markers has a potential to increase with the voxel size increase. This happens due to the increase in the number of voxels that contain partial marker measurements, i.e, measurements of the space that includes both the part of a fiducial marker and a background. At the same time, the number of voxels that are fully within the space corresponding to the markers is decreasing. This phenomenon can be easily illustrated by the definition of the voxel intensity given in Fitzpatrick *et al.* [57]: The intensity value of a voxel can be approximated to the first order, ignoring noise and reconstruction artifacts, by a linear function $\alpha v_i + a_0$, where v_i is the volume of intersection between the marker and the voxel, and α and a_0 are an imaging modality dependent coefficient and the intensity of an empty voxel (air), respectively. For example, for the CBCTs of the RSVP with the voxel sizes of 2 and 4 mm, the measurements of the phantom holder

[1] This definition can be reversed depending on the purposes of the segmentation.

[2] The VOI in this example contains a volume corresponding to the grid points $\{\check{p} \mid -50 \text{ mm} \leq \check{p}_i \leq 50 \text{ mm}\}$, where $i = \{x, y, z\}$. More comments about the VOI selection are provided in Section 3.1.1.

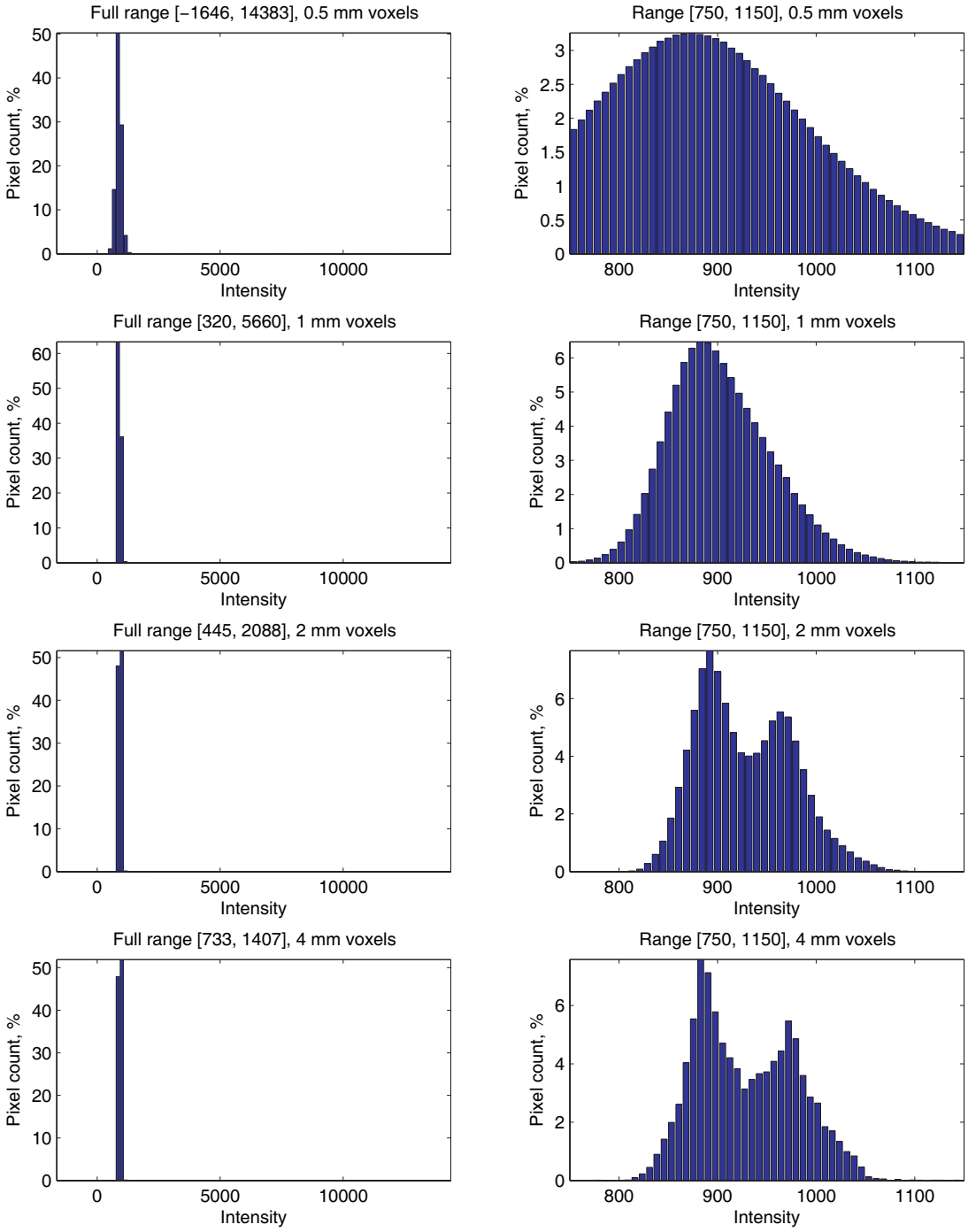


Figure 3.1: Intensity distributions of volumes of interest containing markers in four CBCT images with the voxel sizes of 0.5, 1, 2, and 4 mm reconstructed from the same set of CBCT projections (*left column*) with zoom-ins into [750, 1,150] range of intensity units (*right column*). Different ranges make it challenging to select an intensity threshold value to successfully segment the 3D markers in all images.

and the 3×0.8 mm markers possess similar intensities, and it is not possible to select a threshold to separate the two. For patients, a similar problem can arise with bones. Therefore, an application of a fixed intensity threshold may not be a reliable technique to segment the 3D markers in the CBCT images.

To overcome the aforementioned issues of CBCT scaling with resulting inability to use a single intensity threshold to separate marker and non-marker voxels, we propose a technique for automatic 3D marker segmentation that:

- is independent of the CBCT resolution and scaling;
- selects high-intensity voxels and conducts analysis to reject non-marker measurements based on the known parameters of the marker size and number, not on the intensity values.

We introduce a concept of a marker set to model the markers in a 3D image $\mathcal{V}$. A *3D marker set* is defined as a set of image points in 3D that corresponds to a single marker, and the points are characterized by high intensity values and spatial proximity within one marker set. We will use the term both in the continuous and discrete cases. In a continuous image $\mathcal{V} : \mathbb{R}^3 \to \mathbb{R}$, a 3D marker set is a set of points $p \in \mathbb{R}^3$, while in a discrete array of intensity values $\mathtt{V}$ (data array, or original 3D CBCT image), a 3D marker set is composed of a number of grid points[3] $\check{p}^i \in \Gamma^3$. Hence, we aim to split a grid of the CBCT image, Γ^3, into the subsets M_k, $k = 0, \ldots, n$, such that the sets $\mathrm{M}_1, \ldots, \mathrm{M}_n$ are the 3D marker sets, and M_0 contains all other voxels (measurements of other features and the background). This is a partition of Γ^3, i.e., $\bigcup_{k=0}^{n} \mathrm{M}_k = \Gamma^3$, and $\mathrm{M}_k \cap \mathrm{M}_j = \emptyset$ if $k \neq j$.

As will be discussed later in this chapter, in some cases it may be beneficial to use another construct, the *3D marker models*.[4] The marker model $\mathrm{M}_k^{\mathrm{mod}}$ is a set of points $p \in \mathbb{R}^3$ that compose an ideal marker shape, either a cylindrical or a spherical solid, depending on the fiducial markers used, of known dimensions. The substitution of the marker sets by the marker models eliminates the problem of inevitable quantization of the measurements in a discrete image, and streak artifacts introduced during CBCT image reconstruction. Therefore, such an approach has the potential to improve the outcome of the 2D-3D image registration performed to localize markers in fluoroscopic images (block $\langle\!\langle \mathbf{L} \rangle\!\rangle$ introduced in Flowchart 2.1, which is described in Chapter 5). To the best of our knowledge, such a method of computing and using the 3D marker models has not been previously presented. Currently, the application of the proposed approach has some limitations, such as when the size of the voxels

[3] Recall that the grid point $\check{p}^i$ is the geometrical centre of a voxel i.

[4] A similar concept, called a 3D marker constellation model, is proposed by Fledelius *et al.* [59]. However, it is computed in a substantially different way, by combining the 2D markers segmented in CBCT projections.

is comparable to the size of the fiducial markers, and, consequently, the orientations of the cylindrical markers can not be recovered reliably. However, it may become the state-of-the-art in the future due to a constant improvement in technology, when higher resolution images will become a part of a standard clinical practice.

Flowchart 3.1 provides a detailed scheme for the marker segmentation and modelling block $\langle\langle \mathbf{S} \rangle\rangle$ introduced in Flowchart 2.1. The candidate marker sets $\hat{\mathrm{M}}_k$ are found by successive application of the novel model-based thresholding $\langle\langle \mathbf{S}_1 \rangle\rangle$, clustering $\langle\langle \mathbf{S}_2 \rangle\rangle$, and marker set selection $\langle\langle \mathbf{S}_3 \rangle\rangle$ performed using knowledge about the number and the dimensions of the fiducial markers. These three steps are described in Section 3.1. Section 3.2 provides an explanation of how the 3D marker positions, c^k_{3D}, are computed from the candidate marker sets. Points c^k_{3D} denote the positions of the fiducial markers in the XYZ coordinate system, and are used to quantify the accuracy of 3D marker localization. Next, the correction of the marker set selection is performed, if necessary, at step $\langle\langle \mathbf{S}_4 \rangle\rangle$ as described in Section 3.3, and the 3D marker sets M_k are selected from the candidates $\hat{\mathrm{M}}_k$. Finally, depending on the choice of a template image generation method used in 2D-3D image registration that is later performed in block $\langle\langle \mathbf{L} \rangle\rangle$, a computation of the marker models may be required. This step, referred to as step $\langle\langle \mathbf{S}_5 \rangle\rangle$, is described in Section 3.4. This chapter is concluded with the discussion of the results in Section 3.5.

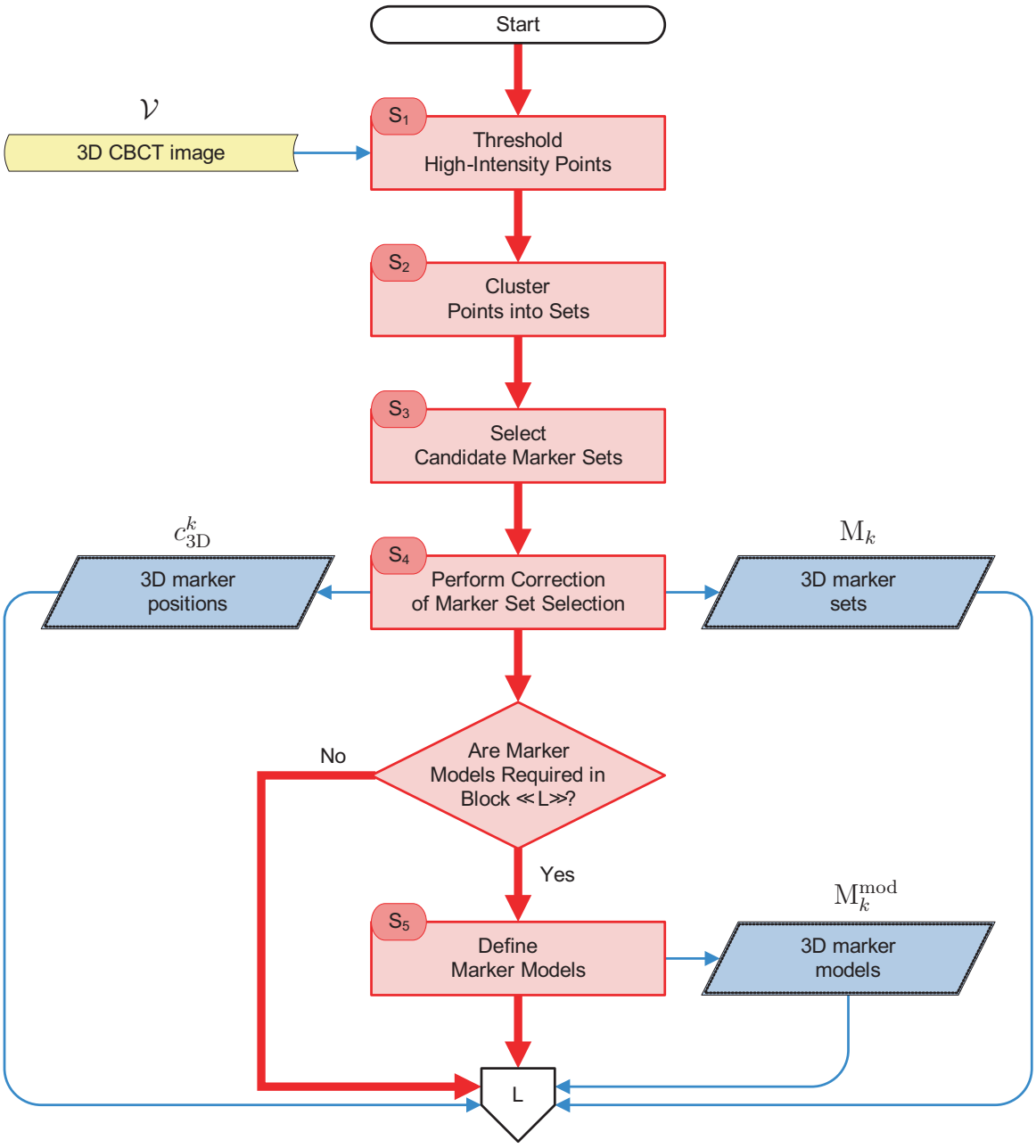
3.1 Marker Set Segmentation in 3D Image

This section presents the methodology of selection of the marker points in the 3D image, and the process of grouping them together in the marker sets that represent markers. As mentioned above, a frequently used technique to segment the markers in the CT or CBCT images is global intensity thresholding [7, 204, 141, 161, 152, 105, 137]. The threshold value θ is usually selected manually by the user or is based on an image histogram, such as proposed by Moseley *et al.* [151, 152], where the threshold is defined as three times the standard deviation of the intensity values. Alternatively, Koch *et al.* [105] and Mahnken *et al.* [137] compute it as:

$$\theta = \gamma \max(\mathcal{V}_{\mathrm{VOI}}), \tag{3.1.1}$$

where $\mathcal{V}_{\mathrm{VOI}}$ is a volume of interest[5] (VOI) in the image $\mathcal{V}$. The value of $\gamma = 0.2$ is set in [137], and the value of $0 < \gamma < 1$ is selected by the operator in [105]. Also, in [137], the value of the computed θ is not allowed to fall below the 3,000 HU CT intensity,

[5]Note that the terms VOI and ROI can be used in reference to discrete and continuous 3D and 2D images, as well as sets and grids on which the images are defined. Selection of a VOI or ROI in the image implies the selection of a VOI or ROI in their corresponding set or grid, and vice versa.



Flowchart 3.1: Block ⟪**S**⟫: Marker segmentation and modelling in 3D.

in order to exclude bones from the thresholding. Due to possibly different scaling and resolution of the CBCT images discussed above, we will design a model-based thresholding that selects high-intensity voxels based on the preliminary knowledge of marker size and number rather than relying on their intensities.

The overall technique of segmenting the markers described in steps $\langle\langle \mathbf{S}_1 \rangle\rangle$ – $\langle\langle \mathbf{S}_3 \rangle\rangle$ is to some extent similar to the one presented by Koch *et al.* [105], though it possesses a number of significant differences. The purpose of marker segmentation in [105] was to provide automatic means for daily setup verification through landmark-based registration of the pretreatment CT and daily megavoltage cone-beam (MVCB) images. While the marker positions in the CT were determined manually, the high-intensity points in the VOI of the daily MVCB were thresholded as defined in Equation (3.1.1). In contrast, we apply the model-based threshold described in Section 3.1.1. Following thresholding in [105], the adjacent points were grouped into contiguous features by the connected component analysis, and the features whose volume exceeded a maximum expected volume of a marker were removed. In order to reduce the computational burden of registering the obtained points to the reference CT points, the former were further grouped into possibly overlapping sets based on close geometrical proximity, and their centroids were used for registration. Another possibility of grouping the thresholded high-intensity points into sets corresponding to the markers based on geometric proximity is proposed by Moseley *et al.* [152], where the points are assigned to the sets based on their CBCT slice location (along Y axis). In our method, we choose not to rely on the contiguity of high-intensity features but only on the proximity of the points to each other due to the following reasons:

- As empirically observed in our testing data sets, it is possible that some reconstruction scenarios produce images with low-intensity voxels within the overall high-intensity markers. This puts the latter in the danger of not being necessarily contiguous, as is demonstrated in Section 3.1.3;
- Relaxation of contiguity requirement also reduces the dependance of the final segmentation result on the threshold value, thus making the method more reliable.[6]

Also, in our method, the candidate marker sets are selected among the high-intensity features based on their size (largest dimension) rather than the volume, which can be a more reliable approach in the presence of reconstruction artifacts in the CBCT images as further discussed in Section 3.1.3.

[6] I.e., if the threshold happens to be higher than some voxels in the middle of the marker, they will still be assigned to the same marker.

3.1.1 Thresholding

At the first step of the segmentation procedure $\langle\!\langle \mathbf{S}_1 \rangle\!\rangle$, a set of high-intensity points Θ is created. The thresholding is based on the assumption that the marker set points are among the highest-intensity points presented in the image: As explained in Section 2.1.1, the radio-opaque fiducial markers are made of the material characterized by high x-ray attenuation properties, where high attenuation translates into high values in $\mathcal{V}$. In many cases, to reduce computational cost and improve marker segmentation or detection, a manually selected volume of interest $\mathcal{V}_{\text{VOI}}$ is analyzed instead of the whole image $\mathcal{V}$, e.g., [105, 7]. The set Θ is computed as:

$$\Theta = \{p \in \Omega_{\text{VOI}} \mid \mathcal{V}_{\text{VOI}}(p) \geq \theta\}, \tag{3.1.2}$$

where $\Omega_{\text{VOI}} \subset \Omega \in \mathbb{R}^3$ is a VOI, $\mathcal{V} : \Omega \to \mathbb{R}$ (for image definition, see Section 2.4.1), and θ is a model-based intensity threshold. The following paragraphs discuss the choice of Ω_{VOI} and θ.

▲ Ω_{VOI}, **Volume of interest**[7]

When the patient is in the treatment position, including the time of the CBCT image acquisition, the target, together with the fiducial markers, is supposed to be in, or in the close proximity to, the LINAC's isocentre. Hence, a rectangular VOI encompassing the XYZ origin can be used:

$$\Omega_{\text{VOI}} = \left\{p \in \Omega \mid p_j^{\min} \leq p_j \leq p_j^{\max},\ j = \{x, y, z\}\right\}, \tag{3.1.3}$$

where $p^{\min}$ and $p^{\max}$ are the vectors containing the lower and upper bounds of the VOI, respectively, in X, Y, and Z dimensions. To include the isocentre, it must hold that $p_j^{\min} \leq 0$ and $p_j^{\max} \geq 0$. The bounds can be determined based on the knowledge of the typical PTV size. Moreover, if the PTV contours are available from the planning CT image, and are expressed as a set of points P in the XYZ coordinate system, and the parameters of the daily couch correction, $t \in \mathbb{R}^3$, are known, we propose to compute the bounds as[8]:

$$p_j^{\min} = \min\left\{p_j^i \mid p^i \in \mathrm{P}\right\} + t_j \tag{3.1.4}$$

and

$$p_j^{\max} = \max\left\{p_j^i \mid p^i \in \mathrm{P}\right\} + t_j, \tag{3.1.5}$$

[7]The paragraphs with parameter description start with ▲ and end with ■ for easy identification in the text.

[8]Note that while in general we use the function notation $\min(\cdot)$, $\max(\cdot)$ with round brackets, where "$\cdot$" can be a set, a sequence or an array, the shorter notation $\min\{\cdot\}$ will be used instead of $\min(\{\cdot\})$ for clarity when "$\cdot$" is a description of the set such as in Equations (3.1.4) and (3.1.5).

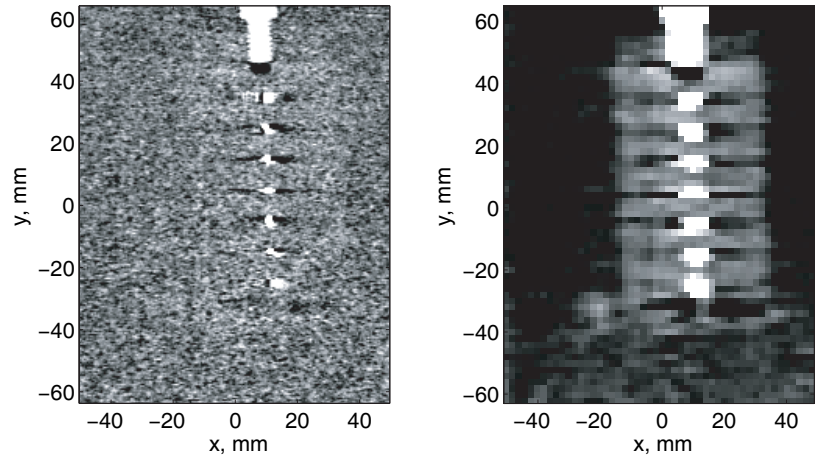


Figure 3.2: CBCT slices of the RSVP with different voxel sizes reconstructed from the same set of CBCT projections. The markers in the image with 0.5 mm voxels (*left*) tend to be smaller than the ones in the 2 mm voxel CBCT (*right*).

where $j = \{x, y, z\}$. The use of the rectangular VOI is favoured for the ease of implementation.

For consistency in all our experiments, we have used a cuboid VOI with $p_j^{\min} = -50$ mm and $p_j^{\max} = 50$ mm. This VOI guarantees the inclusion of all markers for all experimental phantom positions and by far encompasses the markers in the patient images. It also includes other high-intensity measurements, such as of the phantom holder, or the patient bones, allowing to test the reliability of the proposed segmentation procedure. ■

▲ θ, **Segmentation threshold**

To address the weaknesses of the previous approaches in application to the CBCT images associated with the fixed value or image histogram based thresholds, we propose a novel model-based threshold that relies on prior knowledge of the number and size of the fiducial markers. The idea is to select the points of the highest intensities in $\mathcal{V}$ that, combined based on their geometric proximity, may create high-intensity features, the compound volume of which is similar to the compound volume of all fiducial markers, the measurements of which are expected to be observed in the image. The minimal intensity value found in these features is the threshold θ. A valid mechanism to achieve this is by marking iterative change to θ, and comparing between the volume of the thresholded features and the expected volume of the markers.

While for continuous images there arises a non-trivial question of how the features are created from the high-intensity points, and how their volumes are measured, this idea is rather straightforward for algorithmic implementation with discrete im-

ages: Rather than creating and measuring the volumes, we compute the number of voxels, n_{vx}, that are expected to constitute the markers, then sort the intensities in the VOI of the discrete image $\mathtt{V}_{\text{VOI}}$, and finally, segment n_{vx} brightest voxels. The value of n_{vx} is computed in the following way: Let n be a number of the fiducial markers, and δ_{vx} be a voxel size.[9] For cylindrical markers, which we assume all to be of the same size, let r be a radius, and h be a height of a cylinder. Then, the number of the voxels constituting the markers can be computed as a volume of the markers divided by a volume of a voxel, and multiplied by the number of the markers: $n\,\lceil \pi\, r^2\, h/\delta_{\text{vx}}^3 \rceil$, where $\lceil\cdot\rceil$ is rounding towards $+\infty$. The smaller δ_{vx}, the more precise this estimation would be. In reality, for the fiducial markers of 3 to 5 mm long and the CBCT voxels of 0.5 or 1 mm that are often used clinically, it becomes necessary to increase the number of voxels per marker, both to include more voxels containing partial marker measurements into the set Θ, and because the sizes of the markers tend to increase with the increasing voxel sizes. This can easily be illustrated by comparing the CBCT slices of the RSVP in Figure 3.2. Both CBCTs were reconstructed from the same set of the CBCT projections. However, the voxel sizes are different, being 0.5 and 2 mm, respectively. Therefore, we set:

$$n_{\text{vx}} = n \left\lceil \frac{\pi\,(r+\epsilon_1)^2\,(h+\epsilon_2)}{\delta_{\text{vx}}^3} \right\rceil, \tag{3.1.6}$$

where the parameters ϵ_1 and ϵ_2 are selected so that they extend the marker volume uniformly, i.e., $\epsilon_2 = 2\,\epsilon_1$. As explained above, they should also depend on the voxel size. Thus, we set $\epsilon_1 = \kappa\,\delta_{\text{vx}}$ and $\epsilon_2 = 2\,\kappa\,\delta_{\text{vx}}$. The threshold flexibility parameter $\kappa = 1$ is selected based on the experiments with the CBCT images in the phantom and patient testing data sets presented in Section 7.1.2.

Although the inclusion of more voxels by having $\epsilon_1, \epsilon_2 > 0$ often leads to thresholding of features other than the markers, those should be easy to detect based on their size and position. The main task of the thresholding step $\langle\!\langle \mathbf{S}_1 \rangle\!\rangle$ is to select enough points within the markers, regardless of whether or not other high-intensity non-marker measurements are segmented along the way. Those can be removed during the following steps of the segmentation procedure, $\langle\!\langle \mathbf{S}_3 \rangle\!\rangle$ and $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$.

For spherical markers with radii $r_1, \ldots, r_n$,

$$n_{\text{vx}} = \sum_{i=1}^{n} \left\lceil \frac{4}{3}\pi \frac{(r_i+\epsilon_i)^3}{\delta_{\text{vx}}^3} \right\rceil. \tag{3.1.7}$$

We have applied the segmentation technique described in this chapter to a testing data set (not documented in this thesis) with spherical markers 7.83 and 12.65 mm

[9] CBCT voxels are cubic. For anisotropic voxels, it is necessary to account for different values in different dimensions: $(\delta_{\text{vx}})_x$, $(\delta_{\text{vx}})_y$, and $(\delta_{\text{vx}})_z$.

in diameter. As these markers were of a significant size relatively to 1 mm voxels of the CBCT, we used $\epsilon_i = 0$.

After n_{vx} is computed, the values in $\mathtt{V}_{\text{VOI}}$ are sorted in a descending order to obtain a sequence $v_1 \geq v_2 \geq \ldots \geq v_{n_{\text{vx}}} \geq \ldots$. Then, the threshold value is:

$$\theta = v_{n_{\text{vx}}}, \tag{3.1.8}$$

and the set of high-intensity points Θ is computed according to Equation (3.1.2). ■

Naturally, the presence of a large high-intensity non-marker feature in the image can jeopardize thresholding of the marker voxels. This scenario is discussed in Section 3.3.

3.1.2 Clustering

At step $\langle\!\langle \mathbf{S}_2 \rangle\!\rangle$, the set Θ is partitioned, or clustered, into q subsets $\hat{\text{M}}_1, \ldots, \hat{\text{M}}_q$ based on the proximity between the points $p \in \Theta$. Although the number of the markers is known, there is no guarantee that $\mathcal{V}_{\text{VOI}}$ does not contain high-intensity points that belong to the measurements of the features other than the fiducial markers, i.e., a number of subsets that Θ should be partitioned into is not known a priori. This leads us to the use of *hierarchical clustering* method [139] that is based on a concept of the *shortest distance* between sets $\text{A}, \text{B} \subset \mathbb{R}^d$ defined as:

$$\text{dst}(\text{A}, \text{B}) = \inf \{ \|p - s\| \mid p \in \text{A},\ s \in \text{B} \}. \tag{3.1.9}$$

Clustering starts with initialization, where each of the subsets $\hat{\text{M}}_1, \ldots, \hat{\text{M}}_q$ contains only one point. Then, the sets are iteratively merged, starting from the two sets with the smallest shortest distance, until $\min\{\text{dst}(\cdot, \cdot)\}$ becomes larger than η, or the number of clusters reaches $q = 1$.

▲ η, **Minimal distance between the marker sets**

This parameter depends on what is known about the size of the fiducial markers and their relative geometry. The parameter η should be large enough to collect the points of one marker into one set, yet small enough not to include the points of the neighbouring measurements. In all our testing phantom and patient data sets, the distance between the markers was larger than their size, where the *marker size* δ_{m} is defined as the largest of its dimension, i.e., height for cylindrical, $\delta_{\text{m}} = h$, and diameter for spherical markers, $\delta_{\text{m}} = 2r$. Thus, we assign:

$$\eta = \max\left(\delta_{\text{m}}, 1.8\,\delta_{\text{vx}}\right). \tag{3.1.10}$$

The logic behind using δ_{m} in η is that even if only two points per marker set are detected that geometrically correspond to the far ends of the fiducial marker, they

will nevertheless be assigned to one marker. Selecting a maximal value between δ_{m} and 1.8 of the voxel size serves as a safeguard: For coarse resolutions, when δ_{m} is smaller than the voxel size, the neighbouring high-intensity voxels containing partial marker measurements should still be joined into one set. In particular, this situation happens in one of our phantom testing data sets that contains 4 mm voxels and the measurements of the 3 mm long fiducial markers. The value of $1.8\,\delta_{\mathrm{vx}}$ is selected so as to be larger than the distance between the voxels touching diagonally (at their corners), namely $\sqrt{3}\,\delta_{\mathrm{vx}}$, but smaller than the distance to the next nearest voxel, namely $2\,\delta_{\mathrm{vx}}$. If the fiducial markers are placed very close to each other, the value of η should be reduced correspondingly. ■

3.1.3 Selection of Candidate Marker Sets

After the clustering has been performed, the segmentation procedure moves on to step $\langle\!\langle \mathbf{S}_3 \rangle\!\rangle$, where the *candidate marker sets* are selected among $\hat{\mathrm{M}}_1, \ldots, \hat{\mathrm{M}}_q$ by analyzing their sizes. The size of a set $\mathrm{A} \in \mathbb{R}^d$ is measured by its *diameter*, which is defined as:

$$\operatorname{diam}(\mathrm{A}) = \sup\left\{\|p - s\| \mid p, s \in \mathrm{A}\right\}. \tag{3.1.11}$$

For each accepted candidate marker set $\hat{\mathrm{M}}_k$, we require that:

$$\delta_{\mathrm{m}} - \epsilon_{\ell} \leq \operatorname{diam}(\hat{\mathrm{M}}_k) \leq \delta_{\mathrm{m}} + \epsilon_{\mathrm{u}}, \tag{3.1.12}$$

where δ_{m} is the marker size, and ϵ_{ℓ} and ϵ_{u} are the lower and upper margins on the marker set size. The margins are necessary as $\operatorname{diam}(\hat{\mathrm{M}}_k)$ cannot be expected to be equal to δ_{m} due to errors introduced during image acquisition, reconstruction, and quantization (discreteness). In addition, the diameters of the candidate marker sets may typically be larger than δ_{m} due to the inclusion of more high-intensity voxels by setting the flexibility parameter $\kappa > 0$ as explained in Section 3.1.1.

Objects with high atomic number, such as gold fiducial markers, strongly attenuate x-rays, and doing so often induces weak signals in their shadows. This problem is typical of all slice based CT systems, including CBCT. As a result, the reconstructed images demonstrate loss of soft-tissue visibility and *streak artifacts*, a false high-intensity signal in the reconstructed image in a form of bright streaks originating from the high-intensity measurement, e.g., see Figures 3.3 and 3.4. Due to the reduction in image quality, which, in particular, can lead to the decrease of the target localization accuracy, this problem is extensively addressed in medical physics and radiation oncology literature. A number of approaches are proposed that mainly rely on some kind of modification of projections by filtering and/or replacing missing information with the modelled or interpolated data that is followed by the reconstruction

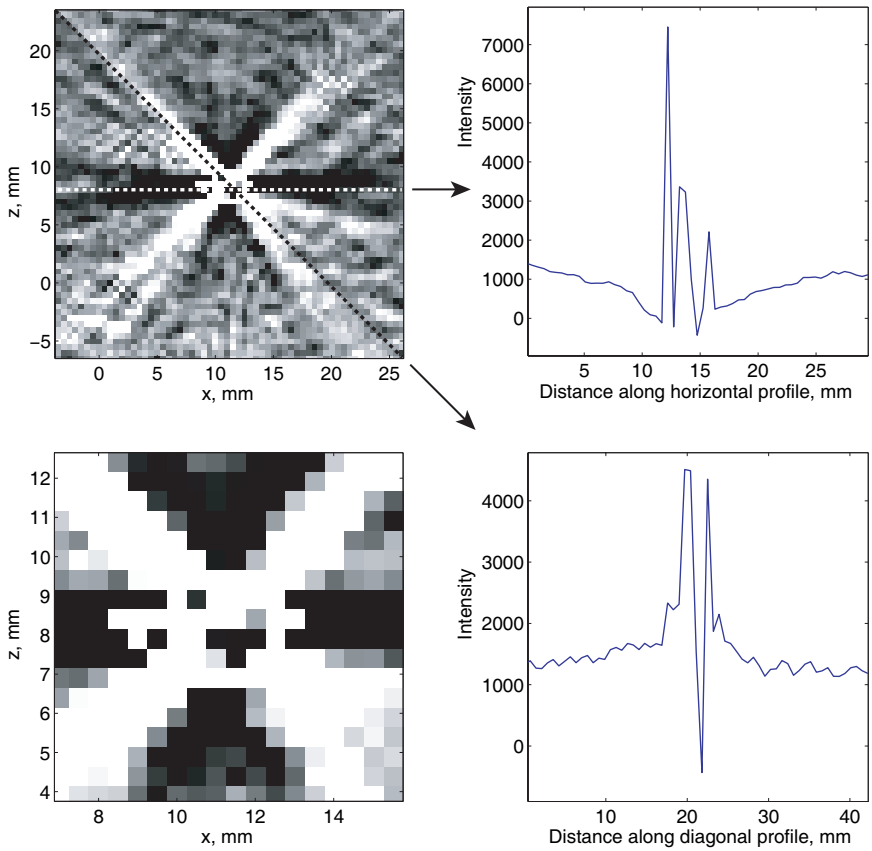


Figure 3.3: A transverse CBCT slice (XZ plane) of the RSVP with 0.5 mm voxels and `short` reconstruction data type (*top left*), its zoom-in (*bottom left*), and its intensity profiles along the horizontal (*top right*) and diagonal (*bottom right*) lines overimposed as the (*dashed lines*) on the (*top left*) image. In addition to the marker in the centre, the image shows severe streak artifacts and the presence of very low-intensity points within high-intensity marker measurement.

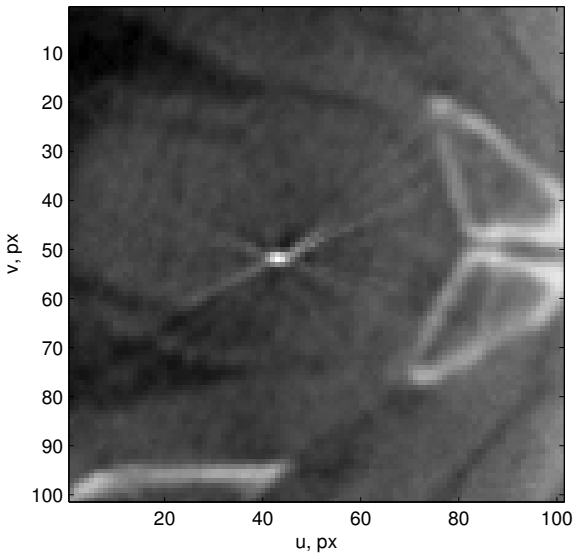


Figure 3.4: A CBCT slice of the prostate patient with streak artifacts around the marker.

of the 3D image with the reduced artifacts, e.g., [151, 18, 245, 137]. As such, these approaches require CT/CBCT projections to perform corrections, and can only be utilized if incorporated into the clinical reconstruction software, as in clinical practice it is not uncommon to delete projections following 3D image reconstruction in order to save space. Therefore, we do not rely on artifact reduction methods.

The XVI software, used clinically for the CBCT reconstruction, allows to choose a number of settings, among which are voxel size and data type, such as `short` and `float`. The data type can influence the quality of the output image significantly in the presence of the fiducial markers. Figures 3.3 and 3.4 show transverse slices of the RSVP CBCT with 0.5 mm voxels and prostate patient with 1 mm voxels reconstructed with `short` data type, respectively. While streak artifacts were a comparable or smaller issue for images with the `float` reconstruction data type, we have not observed low-intensity points inside the high-intensity markers, such as shown in Figure 3.3. We suspect that one of the possible causes for these false low-intensity points may be numerical overflow when using `short` data type. Nevertheless, the images with the `short` data type are included in the further analysis as they are often used clinically.

The severity of the artifacts, and the presence of the low-intensity points inside the markers cause more artifact points to be selected as they can possess intensities similar to the markers and higher than those low-intensity points inside the markers.

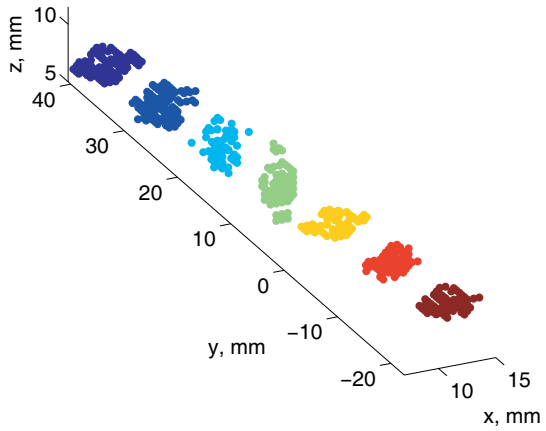


Figure 3.5: Candidate marker sets segmented in the RSVP CBCT image with 0.5 mm voxels and `short` data type. In addition to true marker points, multiple voxels belonging to the streak artifacts were also segmented, thus creating sets of non-regular shapes with increased diameters.

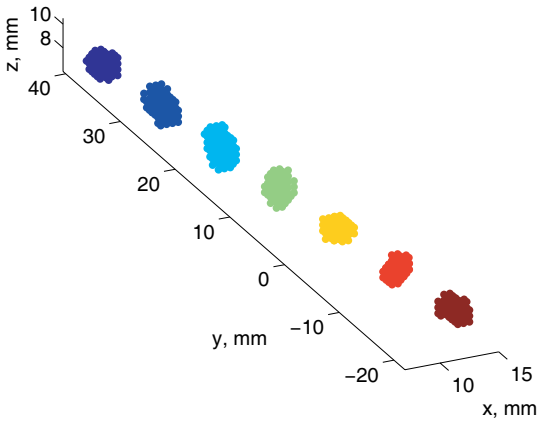


Figure 3.6: Candidate marker sets segmented in the RSVP CBCT image with 0.5 mm voxels and `float` data type better correspond to the shapes of the fiducial markers than those in Figure 3.5.

This is demonstrated in Figures 3.5 and 3.6, where the candidate marker sets are segmented in two RSVP CBCT images with 0.5 mm voxels, reconstructed with `short` and `float` data types, respectively. In comparison to Figure 3.6, the sets in Figure 3.5 acquire a lot of points from the streak artifacts, while some points corresponding to the inner part of the markers are missing, which results in enlarged irregularly shaped sets. This leads us to understanding that the lower and upper marker set size margins, ϵ_ℓ and ϵ_u, should be selected based on information about the overall quality of the CBCT images, including the level of noise and severity of the streak artifacts.

▲ $\epsilon_\ell, \epsilon_\mathrm{u}$, **Marker set size margins**

The values of $\epsilon_\ell = 0.5$ mm and $\epsilon_\mathrm{u} = 11$ mm were selected based on the statistics on the marker set diameters in the phantom and patient data sets. This experiment is described in detail in Section 7.1.3. ■

Note that the size margins can be even larger than the distance between the markers, which is the case for all our phantom data sets, where the distance between the neighbouring marker centres was 10 mm. It still comprises a reliable segmentation scheme, as the points are assigned to one cluster based on another parameter, η, defined in Section 3.1.2. Also, we believe that it is better to have ϵ_ℓ and ϵ_u as loose as the data dictates, even if some non-marker features are included into the candidate

marker sets that can later be removed at the correction step $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$, than to have tighter margins and miss out on the true marker sets.

Appendix D.1 provides implementation details for steps $\langle\!\langle \mathbf{S}_1 \rangle\!\rangle - \langle\!\langle \mathbf{S}_3 \rangle\!\rangle$ of the proposed segmentation procedure, and summarizes them in Algorithm D.1.

3.2 3D Localization of the Fiducial Markers

Marker localization in 3D space refers to the process of finding the geometric location corresponding to the fiducial marker centre in the XYZ coordinate system. In the following, by referring to a *position* of a marker in the CBCT image, or an estimation of the fiducial marker position in the 3D space, we will mean the *centre point* $c^k_{3\mathrm{D}}$ of a candidate marker set $\hat{\mathrm{M}}_k$. In line with accepted practice, we compute this point as the geometric or intensity-weighted centre. In discrete images, where the centres of the segmented voxels (the grid points) are denoted by $\check{p}^i \in \hat{\mathrm{M}}_k$, $i = 1, \ldots, m_k$, the *intensity-weighted centroid* of all voxels belonging to one marker is computed as [57, 228]:

$$c^k_{3\mathrm{D}} = \mathrm{mean}(\hat{\mathrm{M}}_k, w) = \frac{\sum_{i=1}^{m_k} w_i \, \check{p}^i}{\sum_{i=1}^{m_k} w_i}, \quad \check{p}^i \in \hat{\mathrm{M}}_k, \tag{3.2.13}$$

where $w = (w_1, \ldots, w_{m_k})^T$ are the weights computed as $w_i = \mathtt{V}[\check{p}^i] - w_0$, $\mathtt{V}[\check{p}^i]$ is an intensity of the voxel i with the centre in point $\check{p}^i$, and w_0 is the intensity of an empty voxel (air). For the *geometric centroid* [228], all weights are considered to be equal, i.e., $w_1 = \ldots = w_{m_k}$.

The references in [228] suggest that the use of the intensity-weighted centroid provides better marker localization accuracy. The reason behind may be the following: Localization of measurements in a discrete image is characterized with inherent geometric uncertainty [57]. If a voxel contains partial marker measurement, i.e., it includes both a part of a marker and a background, the true position of a high-intensity material within the voxel is not known. However, according to our discrete image model, it is assumed to be concentrated in the geometric centre of a voxel, a grid point $\check{p}^i$. Therefore, using weights based on the voxel intensities helps to reduce the influence of the voxels with the partial marker measurements in the estimation of the marker position.

In the proposed marker localization method, we also compute the centre point $c^k_{3\mathrm{D}}$ as the intensity-weighted centroid defined in Equation (3.2.13) but modify the computation of the weights due to a reason explained below. In our data sets, most of the values $\mathtt{V}_i$ of the marker voxels can be very close in magnitude, i.e., $w_1 \approx \ldots \approx w_{m_k}$, which means the application of Equation (3.2.13) will produce a point very close to the geometric centroid. Therefore, to have an influence, the intensities $\mathtt{V}_i$ have to

be modified. We rescale them so that the values $\mathtt{V}_i$ of the points belonging to one candidate marker set take the range $[0.1, 1]$. The lower bound is selected to be 0.1, and not 0, to give some small weight to the minimal intensity point in a candidate marker set as well.

In some cases, the use of the geometric centroid may be preferable. For example, in the experiment with the spherical markers of 7.83 and 12.65 mm in diameter (not documented in this thesis), we used the geometric centroid to estimate the marker positions. Firstly, the spheres are large in relation to the voxel size. Secondly, the material of some of the spheres was inhomogeneous, and therefore, a simple geometric centroid provided a better localization accuracy.

3.3 Correction for the Marker Set Selection

After the removal of the clusters that fall outside of the accepted size range at step $\langle\!\langle \mathbf{S}_3 \rangle\!\rangle$, the candidate marker sets $\hat{\mathrm{M}}_1, \ldots, \hat{\mathrm{M}}_{\hat{n}}$ are selected. If the number of the sets, $\hat{n}$, is different from the known number of the fiducial markers, n, a correction for the marker set selection is required. This step of the segmentation procedure is referred to as $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$.

If $\hat{n} > n$, some of the sets correspond to non-marker measurements, and have to be removed. Before deciding on the procedure, let us look at several examples from our phantom and patient data sets:

1. In addition to the marker measurements, Ω_{VOI} included a partial measurement of the RSVP insert holder that contained some high-intensity points as demonstrated in Figure 3.7. The measurement was similar in size to the markers.

2. The RSVP had three slightly smaller fiducial markers on the surface in addition to the seven internal fiducial markers as shown in Figure 3.8. If Ω_{VOI} is large enough to include those, the smaller markers will also be accepted as the candidate marker sets.

3. Depending on the choice of Ω_{VOI} in the patient images, the candidate marker sets can include high-intensity non-marker measurements, such as those of the bones, as demonstrated in Figure 3.9.

What we observe from these examples, is that the non-marker sets of high-intensity points are often located further away from the target denoted by the fiducial markers. Hence, it can be argued that all situations described above could be eliminated with a tighter choice of Ω_{VOI}. However, we believe that a method of automatic processing should not rely on a perfect choice of a parameter. An attempt to

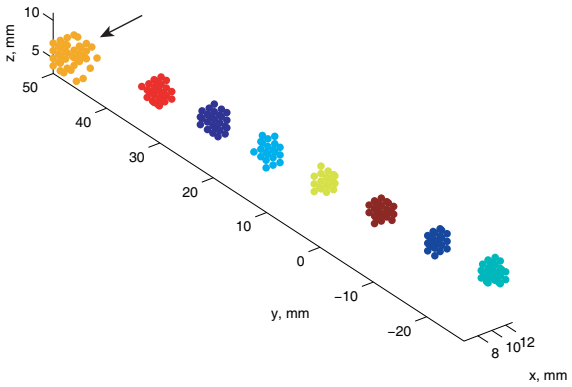


Figure 3.7: In the RSVP image, Ω_{VOI} included a part of an insert holder measurement (*arrow*), that was similar in size to the markers and contained high-intensity points.

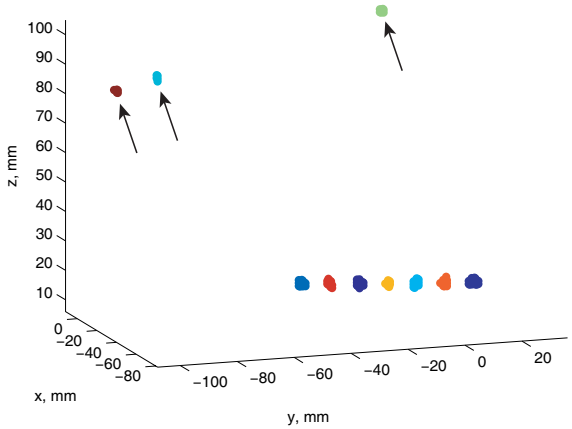


Figure 3.8: In addition to seven internal fiducial markers, the RSVP had three slightly smaller markers on its surface (*arrows*).

tighten the VOI, or a requirement to provide a VOI that does not include any other high-intensity measurements except for those of the fiducial markers, will reduce the generality of the method and may require an additional user interference with the possible alteration in Ω_{VOI} for each and every image, which defeats the purpose of automation. In all of our data sets, which include CBCT images of the prostate cancer patients, and the RSVP and Quasar phantoms, the target, and hence the markers, are located in close proximity to the LINAC's isocentre, the point $p^{\text{prox}} = (0, 0, 0)^T$. Therefore, if $\hat{n} > n$, we remove the candidate marker sets that are furthest away from p^{prox}. We believe this is a reasonable strategy for many scenarios, as the positioning of the target close to the isocentre is a typical geometrical setup for treatment. The implementation details for this correction procedure are provided in Algorithm D.2. In cases in which the point p^{prox} is not available, we suggest that one may seek the sets that are positioned in a certain geometrical relationship known from the clinical setup.

If $\hat{n} < n$, we assume that all found candidates represent true markers, and accept them as the final marker sets $\text{M}_1, \ldots, \text{M}_{\hat{n}}$. One of the possible reasons that not all markers have been identified may be the low value of ϵ_{u}, the upper marker set size margin (Section 3.1.3). If streak artifacts are large enough to violate this size requirement, the sets are removed. Another possible reason is that the voxels belonging to the measurements of the two neighbouring markers were located next to each other, and hence the points of the two measurements were combined into one cluster, and discarded later due to a large size. This can be observed in the RSVP

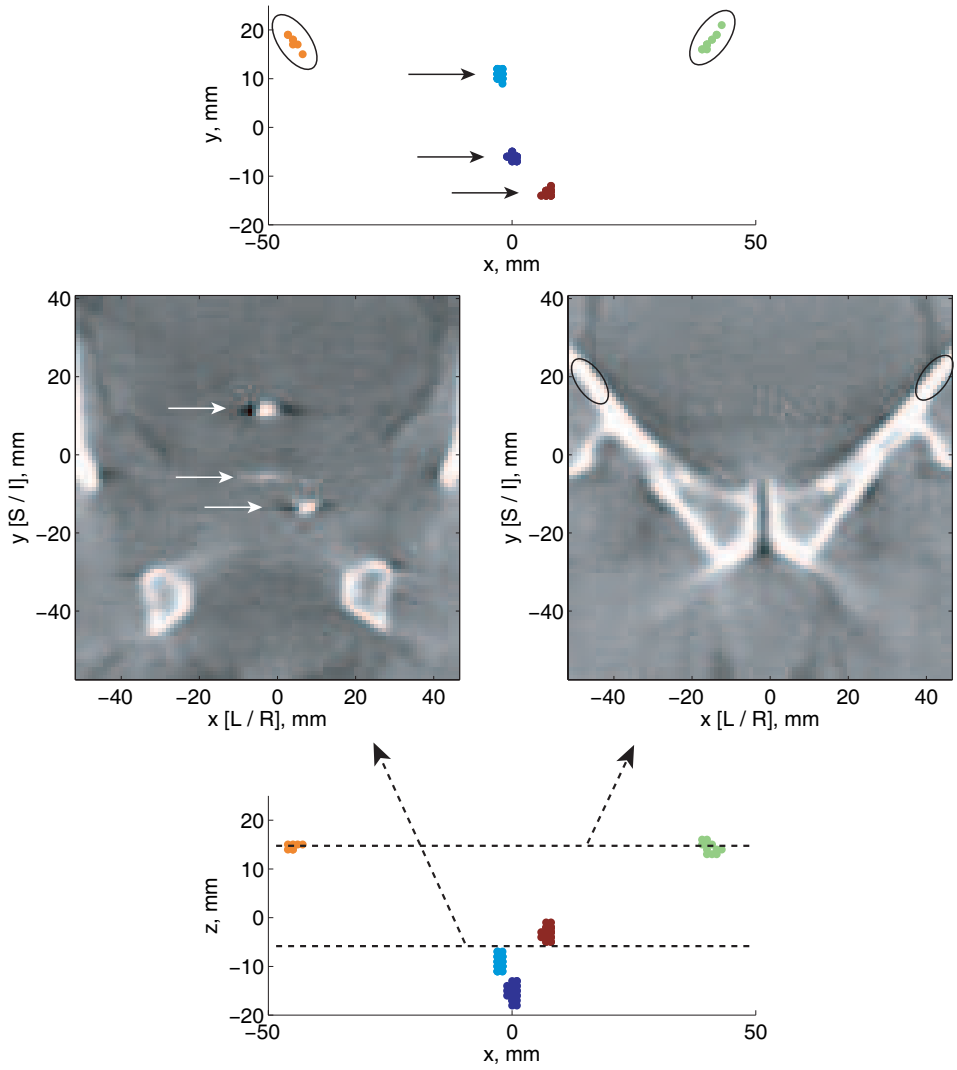


Figure 3.9: As shown in the XY view (*top*) and two coronal CBCT slices of a prostate patient (*XY plane, middle*), that were selected at Z = −6 mm and Z = 15 mm as schematically shown at the *bottom*, the part of the segmentation procedure described in steps $\langle\langle \mathbf{S}_1 \rangle\rangle - \langle\langle \mathbf{S}_3 \rangle\rangle$ selected three sets corresponding to the true markers (*arrows*) and two sets that were parts of the bone measurements (*ellipses*).

images with 4 mm voxels.

The following situations have never happened with our testing data sets, however, it is important to include their description for the generality of the methods. If $\hat{n}$ becomes smaller than n after step $\langle\langle \mathbf{S}_3 \rangle\rangle$, the discarded sets should be analyzed. If the number of the points in any particular one is several times larger than what is expected for a marker (as defined in Equation (3.1.6)), it is possible that although the markers possess reasonable intensity and could successfully be detected, Ω_{VOI} also contained a large higher-intensity measurement that took the point quota away from the markers. An example may be a metallic implant in the patient's body. In this case, the large measurement should be digitally removed, and the whole marker segmentation procedure must be repeated. We do not recommend digital removal of the smaller measurements as those, in fact, may be partially recovered markers. In a case of a very noisy image, where there are many isolated high-intensity points, preliminary spatial smoothing of the image may be helpful.

As presented in Section 7.1.4, the application of the steps $\langle\langle \mathbf{S}_1 \rangle\rangle - \langle\langle \mathbf{S}_4 \rangle\rangle$ to phantom and patient testing data sets resulted in 100% of true positives and 0% of false positives, except for the RSVP images with 4 mm voxels, where the numbers were 83% and 2%, respectively. The numerical results in Section 7.1.4 also demonstrate that it is important to have both $\langle\langle \mathbf{S}_3 \rangle\rangle$ and $\langle\langle \mathbf{S}_4 \rangle\rangle$ steps in the segmentation procedure as they both take part in successful removal of the non-marker sets.

3.4 Marker Models

Markers in the CBCT images, as judged by human observers and resulting from intensity thresholding or edge detection methods, tend to be of a larger size in comparison to the fiducial markers themselves and their measurements in the 2D fluoroscopic images. Typically, this is due to reconstruction artifacts. Another factor is the discreteness of the images, and the fact that the CBCT voxels are often larger in size than the pixels in the fluoroscopic images.[10] This may later cause positioning uncertainty in the process of 2D-3D image registration. One example of a method to overcome this problem is that of Adamson and Wu [7], who proposed to postprocess the measurements of long coil markers obtained by thresholding the CBCTs in the following way: In order to reduce their diameter, the measurements are analyzed slice by slice (perpendicular to the marker's lengthwise direction that is known from the setup) so as to select only a central point in each slice and discard all other points. Naturally, this technique can not be applied to our case, as the markers are significantly shorter

[10] For example, while the CBCT images with 0.5 or 1 mm voxels are often used clinically, the pixel size in our fluoroscopic images measured at the isocentric plane is 0.2604 mm.

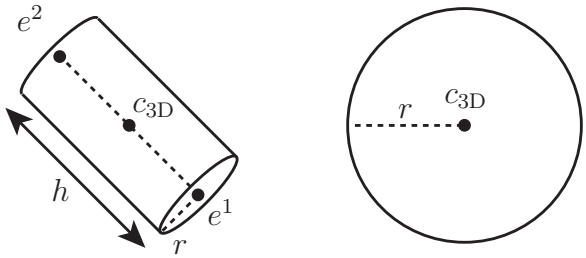


Figure 3.10: Diagrams of the cylindrical (*left*) and spherical (*right*) marker models.

(3×0.8 mm as compared to coils of size $3 - 4$ cm $\times$ 0.35 mm). Furthermore, the markers in our application can have different orientations relative to the CBCT slices.

In this section, we propose a novel technique, that can be used to improve the outcome of 2D-3D image registration. It is referred to as step $\langle\langle \mathbf{S}_5 \rangle\rangle$ in Flowchart 3.1. We propose to utilize marker models instead of the marker sets in the process of registration. The *marker models* are mathematical representations of the markers, such as a continuous cylindrical or spherical model of the known dimensions. If the orientations of the cylindrical fiducial markers can be estimated from the corresponding marker sets with reasonable accuracy, the models can be used instead of the marker sets in the process of 2D-3D image registration between the CBCT and the fluoroscopic images.

To the best of our knowledge, no methods of marker model estimation based on the segmented CT/CBCT voxels have been reported. Several studies use the geometry of the system to estimate the marker positions in 3D from their 2D positions in fluoroscopic or CBCT projection images [7, 170, 141]. An interesting approach was proposed by Fledelius *et al.* [59] for marker segmentation in the CBCT projections: The candidate measurements of the cylindrical markers were segmented in all projections by a blob detection routine, and the traces of the blobs in 2D were recovered. Then, the traces that could not reasonably correspond to an object in 3D were rejected. From the remaining, a 3D constellation model was constructed consisting of the size, position and orientation of the markers. Finally, the model was projected onto the CBCT projection images to serve as a template to yield final reliable segmentation results. Poulsen *et al.* [178] proposed a semi-automated method for constructing a 3D marker model of an arbitrarily shaped marker from its 2D projections. The markers were segmented by intensity thresholding in several CBCT projections with large angular separation, good marker contrast and uniform background. Then, the 3D marker model was constructed by finding the intersection of all the candidate voxels that were computed from each of the segmentations by raytracing with parallel rays. The CBCT voxels and CBCT projection pixels were of equal size.

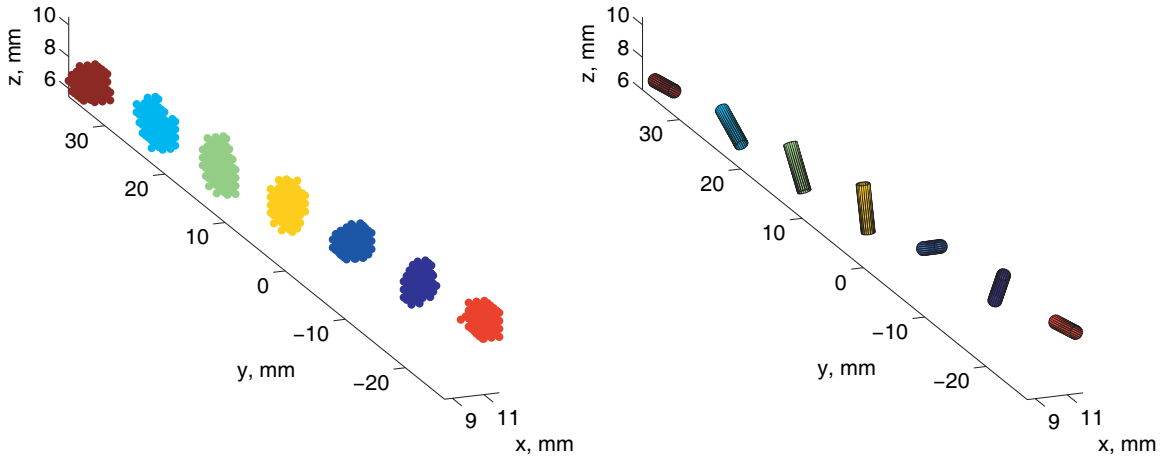


Figure 3.11: Marker sets segmented in a CBCT image of the RSVP with 0.5 mm voxels (*left*) and their corresponding marker models (*right*).

The remainder of this section discusses the proposed method for defining the 3D marker models from the marker sets. As shown in Figure 3.10 (*left*), a cylindrical marker is described by a cylindrical solid of a known radius r enclosed between two planes perpendicular to the cylinder's axis, at distance h (cylinder's height) from each other. Point c_{3D} is a centre of the model and is located on the cylinder's axis equidistantly from the planes. To reveal cylinder orientations from the segmented marker sets, we apply *weighted principal component analysis* (PCA) to the points of each set. In brief, the PCA computes the directions $t^1 \perp \ldots \perp t^d$ of the maximal variance in the data. For the markers whose length is several times larger than their diameter, it is expected that the first principal direction t^1 would provide estimation of the marker orientation for each marker set, i.e., the orientation of t^1 is used to compute the orientation of the line between the endpoints of the cylindrical model e^1 and e^2 (see Figure 3.10, *left*). The use of the weights in the PCA is motivated by the ideas explained in Section 3.2: The assumption is that the brighter the voxel, the larger is the fraction of the marker measurement that is contained in its volume. Hence, the weights w_i for the voxels should be proportional to their intensities $\mathtt{V}_i$, and are computed as explained in Section 3.2 by rescaling $\mathtt{V}_i$ to the range $[0.1, 1]$ for each marker set. Algorithm D.3 provides implementation details for the computation of the marker model orientations. An example of the marker sets and their corresponding cylindrical marker models is demonstrated in Figure 3.11.

A spherical marker can be modelled by a sphere of a known radius r centered at point c_{3D} (see Figure 3.10, *right*). The detailed formulations and derivations of the cylindrical and spherical marker models are provided in Appendix C.3.

3.5 Discussion

In this chapter, we discussed novel and adapted techniques that constitute the proposed marker segmentation and modelling procedure used for marker localization in 3D space. First, to overcome the challenges of selecting an appropriate global intensity threshold for a CBCT image to segment the marker points, discussed in the introduction to this chapter, a novel model-based threshold was proposed (Section 3.1.1). To increase the probability of selecting only the marker points rather than non-marker high-intensity features, and to group the points into features corresponding to the markers, the thresholding was followed by clustering (Section 3.1.2), marker set selection (Section 3.1.3), and correction procedures (Section 3.3), all of which exploited knowledge of marker size and number, and marker geometry in the 3D space. As a result, we obtained the 3D marker sets, which consisted of voxels. To overcome the discrete nature of the 3D marker sets, the use of the 3D marker models was proposed (Section 3.4). The models are continuous mathematical entities describing cylinders or spheres of sizes that correspond to the real fiducial markers. They can be used in place of the segmented marker sets to improve the appearance of the markers in the 2D template image that is formed in the process of 2D-3D image registration between the CBCT and the fluoroscopic image. The computation of the marker models reveals the marker orientation in 3D, which can potentially be used in other applications, such as marker-based 3D-3D image registration or an assessment of the organ deformation over time.

The validation experiments on phantom images, described in Sections 7.1.6 and 7.1.7, suggest that the proposed procedure for marker segmentation and modelling offers sufficient marker localization and orientation accuracy to be used in clinical application. The magnitudes of the marker positioning errors for the CBCT images with 0.25, 0.5, 1, 2, and 4 mm voxels are mostly well below 1 mm, with the exception of the 4 mm voxels data set, where the maximal error was around 2 mm, and the 0.5 mm data set with the `short` data type, which has large streak artifacts, where the maximal error slightly surpassed 1 mm. For most data sets, the positioning accuracy depends almost linearly on the voxel size. For a detailed analysis, see Section 7.1.6.

From the results presented in Section 7.1.7, it follows that the marker models can be successfully computed for the CBCT images with 0.25 and 0.5 mm voxels, with the exception of the images reconstructed with the `short` data type. With some caution, they can also be recovered from most images with 1 mm voxels. If the voxels are larger than 1 mm, or large reconstruction artifacts are present, the marker models should not be used, and the marker localization procedure should proceed by only using the marker sets segmented from the CBCT images. We have seen some correlation between the accuracy of the found marker position and orientation and

the marker orientation in the 3D space. For detailed analysis, see Section 7.1.7.

Some comments on implementation are given in Appendix D.1. The methods described in this chapter are summarized in Algorithms D.1 – D.3.

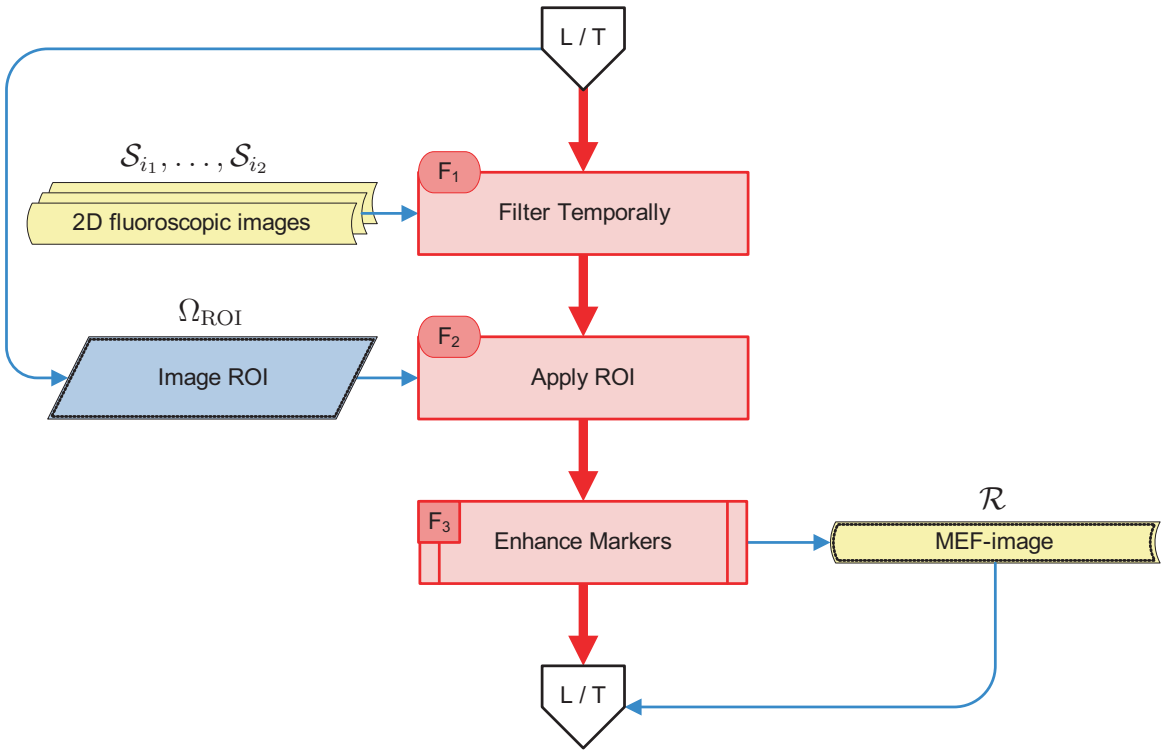
Chapter 4

Design and Adaptation of Filters for 2D Images

This chapter describes the filtering procedures performed on the 2D fluoroscopic images. Denoted as block $\langle\!\langle \mathbf{F} \rangle\!\rangle$ in Flowchart 2.1, it is called from both the marker localization block $\langle\!\langle \mathbf{L} \rangle\!\rangle$ and the tracking block $\langle\!\langle \mathbf{T} \rangle\!\rangle$. Therefore, we have decided to discuss it before the respective descriptions of the blocks $\langle\!\langle \mathbf{L} \rangle\!\rangle$ and $\langle\!\langle \mathbf{T} \rangle\!\rangle$ provided in Chapters 5 and 6. As shown in Flowchart 4.1, block $\langle\!\langle \mathbf{F} \rangle\!\rangle$ consists of three subblocks, *temporal filtering*, that operates on fluoroscopic images, $\mathcal{S}_{i_1}, \ldots, \mathcal{S}_{i_2}$, application of the *regions of interest* (ROIs), Ω_{ROI}, computed in $\langle\!\langle \mathbf{L} \rangle\!\rangle$ and $\langle\!\langle \mathbf{T} \rangle\!\rangle$, and a novel *marker enhancement filter* (MEF). These actions produce an image $\mathcal{R}$ that is ready for image registration in either block $\langle\!\langle \mathbf{L} \rangle\!\rangle$ or block $\langle\!\langle \mathbf{T} \rangle\!\rangle$. The temporal filtering reduces the noise, while the MEF amplifies the markers, suppresses non-marker features, and flattens the background. The application of the MEF to Ω_{ROI}, rather than the whole image, helps to reduce the computational burden and increase the reliability of the filtering procedures.

The goal of the temporal filtering, referred to as step $\langle\!\langle \mathbf{F}_1 \rangle\!\rangle$ in Flowchart 4.1, is to reduce the impact of noise. In order to select a filter and its parameters, several commonly used digital filters are reviewed in Section 4.1.1, and the results of their application to image sequences are analyzed in Section 4.1.3 with respect to image quality, measured by a contrast-to-noise ratio (CNR). The CNR, defined in Section 4.1.2, is sensitive to both changes in noise levels and blurring due to the degradation of the temporal resolution. As the filtering procedure is a preprocessing step for the application of the MEF, we aim to achieve a better CNR in the filtered images compared to the original data by using a simple and computationally efficient method.

The application of the ROI, denoted as step $\langle\!\langle \mathbf{F}_2 \rangle\!\rangle$ in Flowchart 4.1, is a straightforward procedure, and as such, does not require much discussion. This step can be



Flowchart 4.1: Block $\langle\!\langle \mathbf{F} \rangle\!\rangle$: Temporal and marker enhancement filtering of 2D fluoroscopic images.

performed either before or after the temporal filtering $\langle\!\langle \mathbf{F}_1 \rangle\!\rangle$. In our implementation in Gryphon software, the same image opening procedure is used for multiple tasks, such as image registration and image preview. To allow the preview of the temporally filtered images, for example, in order to visually assess the improvement in quality, temporal filtering is built into the image opening procedure, and is applied to the whole image. Therefore, we reduce the image to the ROI after temporal filtering, and before the application of the MEF. Since the ROIs in both $\langle\!\langle \mathbf{L} \rangle\!\rangle$ and $\langle\!\langle \mathbf{T} \rangle\!\rangle$ blocks are determined before the calls to the filtering block $\langle\!\langle \mathbf{F} \rangle\!\rangle$, and are computed in a similar fashion, we describe the principles of ROI computation in Section 4.3, rather than in Chapters 5 and 6.

Next, the design of a novel filter, the MEF, is presented. It is schematically shown as block $\langle\!\langle \mathbf{F}_3 \rangle\!\rangle$ in Flowchart 4.1. In brief, it enhances the markers by conducting local intensity analysis at a particular scale, i.e., it seeks to enhance features of a certain size. The MEF is developed within the scale-space framework [127], described in Section 4.2.2, which found its application in design of descriptors for scene matching in computer vision (blob detection), e.g., [126, 96, 134, 148, 215], and feature enhancement in medical images, mostly line-like vessels, e.g., [132, 189, 121, 63]. We analyze the advantages and shortcomings of existing methods in Section 4.2 in application to our images, and then present the proposed filter in Section 4.4. To the best of our knowledge, there is no other marker-specific filter in the literature. In addition, our filter is fortified with a novel contrast enhancement function that becomes especially important when the contrast of images is low. The principles of the proposed MEF are not specific to our localization and tracking method, and can be utilized in other marker detection or tracking applications.

4.1 Noise Reduction with Temporal Filtering

All radiographic image guidance techniques, including fluoroscopy, can give a significant imaging dose to a patient [158, 89]. Although it is possible to obtain fluoroscopic images with higher contrast by adjusting the hardware settings of image acquisition, this results in an undesirable increase in x-ray exposure. In the view of the current philosophy for dose management adopted by the diagnostic imaging community, which can be summarized by "as low as reasonably achievable" [158], another option to improve the quality of the images, and make them suitable for automatic processing, is digital image enhancement.

A common approach to noise suppression in the preprocessing of medical images involves the use of simple *spatial filtering*, such as *mean*, *median*, or different types of *linear low-pass filters* [168]. Although spatial filtering can provide a substantial

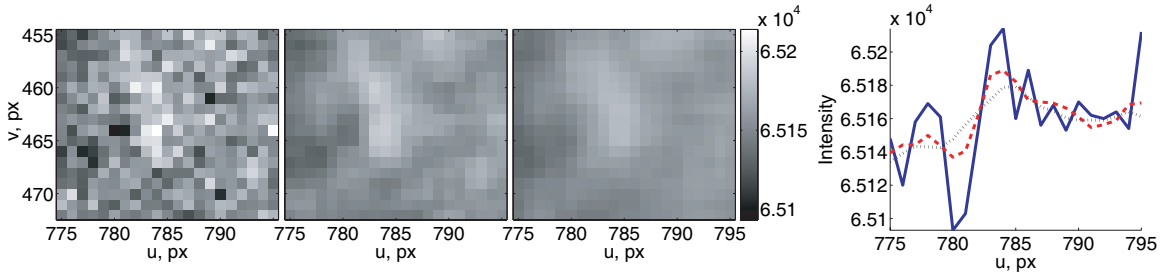


Figure 4.1: The results of the application of the spatial mean filter to a marker extracted from a fluoroscopic image of a prostate patient: the original image (*left image*), after mean filtering with 3×3 pixels kernel (*central image*) and 5×5 pixels kernel (*right image*). The graph at the *far right* shows the intensity profile of the row at the 464 pixels level that contains the highest intensity pixel. The *solid line* denotes the original measurement, *dashed line* the measurement filtered by the 3×3 filter and *dotted line* by the 5×5 filter. Contrast in the filtered images, measured as the absolute difference between the brightest and the darkest intensity values, amounts to $\sim 50\%$ and $\sim 40\%$ of the original contrast for the images processed by the 3×3 and 5×5 filters, respectively.

reduction in the high (spatial) frequency noise, these filters tend to smooth edges of the features present in the image and this causes details smaller than the size of the filter kernel to be suppressed. The size of the fiducial markers that are used for all phantoms and patients in our data sets is 3×0.8 mm, which results in 12×3 pixel features with our current geometric setup, where the size of a pixel side at the isocentric plane (UV coordinate system) is 0.2604 mm. Since effective spatial filters have kernels that are at least on the order of the marker width, spatial filtering often degrades the quality of an image from registration perspective by blurring the markers, as can be seen from examples provided in Figure 4.1. On the other hand, the use of smaller kernels does not provide sufficient noise reduction.

Instead of spatial filtering, we will use *temporal filtering* to improve the contrast of the markers and to decrease the high (temporal) frequency noise. Combining information from multiple frames inevitably raises a question of whether the ultimate goal of tracking the markers to quantify the motion of the target is jeopardized by *motion blurring* and the *reduction in temporal resolution*. In Section 4.1.3, we present the results of the application of several filters to a fluoroscopic image sequence with moving markers, and demonstrate how to select a filter that reduces noise but does not introduce excessive blurring. Concerning the loss in temporal resolution, the main goal of intra-fraction motion tracking is the detection of long term drifts that are more important for margin design than fast incidental moves of the target. In

other words, if the filter improves the quality of the fluoroscopic images enough to sufficiently increase reliability of the marker localization and tracking procedures, but substitutes the initial markers by the "moving average markers",[1] we consider it a worthwhile tradeoff.

We start by reviewing existing temporal filters for fluoroscopy in Section 4.1.1, then proceed with the definition of the contrast-to-noise ratio as a measure of image quality in Section 4.1.2. Finally, we apply the filters to a fluoroscopic sequence showing Quasar phantom with moving markers, compare the results, and select the temporal filter in Section 4.1.3.

4.1.1 Temporal Filters for Fluoroscopy

In this section, we summarize several approaches reported in the literature that apply temporal filtering to fluoroscopic images and discuss their suitability to our marker localization and tracking application. Let us denote the fluoroscopic image sequence as $\mathcal{S}_0, \ldots, \mathcal{S}_{m-1}$, where all images of the sequence are defined on the same ROI. One of the earliest techniques proposed in the literature is a simple form of *recursive filtering* [17, 192]. The output of the filter is the weighted average of a current unfiltered image $\mathcal{S}_i$ and a preceding filtered image $\bar{\mathcal{S}}_{i-1}$:

$$\bar{\mathcal{S}}_i = (1-\lambda)\,\bar{\mathcal{S}}_{i-1} + \lambda\,\mathcal{S}_i, \tag{4.1.1}$$

where $\bar{\mathcal{S}}_0 = \mathcal{S}_0$, and $0 < \lambda \leq 1$ is a weighting factor for a current frame (the value $\lambda = 1$ corresponds to no filtering). The choice of λ depends on the assumptions about the motion observed in the image sequence, where higher λ should correspond to higher speed of motion.

Another approach is the use of non-recursive filters, also known as *finite impulse response filters* (FIR) [80, 32]. In this case, we compute the filtered image as:

$$\bar{\mathcal{S}}_i = \sum_{n=0}^{\ell-1} w_n\,\mathcal{S}_{i+n}, \quad 0 \leq i \leq m-\ell, \tag{4.1.2}$$

where ℓ is a filter kernel size and w is a vector of filter coefficients.[2] All FIR filters considered in this section have symmetric impulse response, i.e., $w_n = w_{\ell-1-n}$, and the coefficients are scaled so that $\sum_{n=0}^{\ell-1} w_n = 1$. One of the simplest filters, the mean

[1] The *moving average* term here is used in a broad sense and refers to the fact that the consecutive filtered images are produced from the intersecting sets of unfiltered frames. For example, if the filtered image k is computed from the frames $i_1, \ldots, i_m$, then the next image $k+1$ can be produced from the frames $i_2, \ldots, i_{m+1}$.

[2] The indices in $\bar{\mathcal{S}}_i$ and $\mathcal{S}_{i+n}$ refer to the sequential numbers of the images in the corresponding sequences rather than time instances, i.e., the filter is defined in a causal manner.

filter (frame averaging, rectangle filter), corresponding to $w_n = 1/\ell$, was reported to be used in fluoroscopic sequences by Adamson and Wu [7]. The mean filter, however, introduces significant blurring when motion is present, which will become obvious later in this chapter, hence we will investigate a number of other FIR filters in Section 4.1.3.

Another method that was used for fluoroscopic images is a filter based on the *Karhunen-Loève transform* (KLT) [51, 227]. The KLT operates by exploiting the second and higher order correlations between neighbouring frames. As noted by Ding *et al.* [51], who used it for cardiac MRI, the KLT may be especially suited for pseudo-periodic signals, such as heart beat or respiration. Although in many cases the motion of the prostate is irregular, small pseudo-periodic motion caused by respiration can sometimes be observed. Therefore, we decided to evaluate this method as well, and compare it to the recursive and FIR filters in Section 4.1.3. The detailed description of the KLT-based filter can be found in [51], but we summarize it in Appendix C.1 for the convenience of the reader.

Finally, a method that results in the *"motion-enhanced" image* was also reported in the literature [24, 50]. The filtered image is created as a difference between an average image (computed as a mean of all frames in the sequence) and the current frame. This filter was used to enhance moving structures in markerless respiratory tracking. As such, it is inapplicable to prostate motion tracking as the prostate may often remain stationary, in which case the filtered image will only contain fluctuations of noise while the markers can be completely removed. Section 4.1.3 presents the results of application of the recursive, FIR, and KLT methods to a testing data set, and quality comparison of the filtered images is performed in terms of the CNR, formulated in the next section, and visual examination.

4.1.2 Image Quality Assessment with Contrast-to-Noise Ratio

A *contrast-to-noise ratio* (CNR) is a measure of contrast between two signals in presence of noise. In our case, the CNR characterizes the visibility of a marker on its local background. Let us denote the mean intensity of the points belonging to the marker by μ_{M}, mean intensity of the background points by μ_{B}, and the standard deviation of the intensities of the background by σ_{B}. Following Siewerdsen *et al.* [202], we will define the CNR to be the signal-on-background contrast divided by a standard deviation of the background as a measure of noise [202]:

$$\mathrm{CNR} = \frac{|\mu_{\mathrm{M}} - \mu_{\mathrm{B}}|}{\sigma_{\mathrm{B}}}. \tag{4.1.3}$$

Note that σ_B is positive by its very nature, as it is not possible to achieve an ideally flat background with the current imaging equipment.

Other definitions of the CNR are also used in the literature. For example, the denominator of the ratio can be a standard deviation of the intensity of the signal points σ_M [207, 227][3] or $\sqrt{\sigma_M^2 + \sigma_B^2}$ [112]. The general idea is to measure a pixel or voxel noise level, which in most cases should be similar for the signal and the background [202]. However, the estimate of this noise is more realistic if it is obtained from regions in the image, in which the intensity is homogeneous. This helps to ensure that the standard deviation is not artificially increased by the intensity variations attributed to the signal rather than noise. Given the relatively small marker size in our application, we conclude that the most accurate estimation of the noise in our case is σ_B obtained from the background points in close proximity to the marker.

The computation of the CNR demands a definition of the points that belong to the marker and to the background, and their intensities. While it is generally easy to select the background points, the determination of the marker points can be more involved. Ideally, the marker is a set of points in the 2D image that includes[4] intensity measurements of the fiducial markers. Precise automatic localization of these points is only possible under idealized conditions: knowledge about the exact fiducial marker location and orientation in the 3D space, ideal system calibration and an accurate model of image formation. It is very unlikely to satisfy these requirements in practice. Even more so, the 2D projection image is discrete, and some of the pixels contain only partial marker measurements, combined with the background. Hence, we define the *points of the marker* in a discrete image as a set of the brightest grid points in some ROI $\Phi_M \subset \Gamma^2$, that is expected to contain all points of one marker and possibly its immediate background. The choice of Φ_M for our testing data set will be explained in Sections 4.1.3.

The principle of automatic selection of the brightest points in Φ_M is similar to the model-based threshold method described in Section 3.1.1 modified for the 2D. Let n_M be a number of points that can constitute one marker. It is estimated in the following way: Let r denote the radius of the cylindrical fiducial marker, h its height, and δ_{px} the pixel size of the fluoroscopic image. When the fiducial markers are inserted into the patient's prostate, every effort is made to place them so that they can be imaged in the same sagittal slice of the daily CBCT (parallel to the YZ plane). This also means that in the lateral fluoroscopic images all markers should appear full-length, or almost full-length. Hence, an area occupied by the marker can

[3] In [227], it is called a signal-to-noise ratio (SNR).

[4] In the projection images, the measurements of all objects that x-rays encounter on their path from the source to the detector, are superimposed. Hence, each point in the image contains measurements of multiple objects.

be approximated by $2rh$, and the number of pixels is then computed as:

$$n_{\mathrm{M}} = \left\lceil \kappa \, \frac{2rh}{\delta_{\mathrm{px}}^2} \right\rceil, \tag{4.1.4}$$

where $\lceil \cdot \rceil$ is rounding towards $+\infty$. Based on the same considerations, the number of pixels for a spherical marker of radius r can be estimated as $n_{\mathrm{M}} = \left\lceil \kappa \, \frac{\pi r^2}{\delta_{\mathrm{px}}^2} \right\rceil$. The scalar κ is used to exclude the marker borderline points that only contain partial marker measurements. These points can be close in intensity to the background points, and using $\kappa = 1$ often results in selection of some background points in lieu of marker points. Our experiments with the phantom and prostate patient fluoroscopic sequences indicated that $\kappa = 0.7$ is a reasonable choice.

After n_{M} is estimated, the mean and standard deviation values of the marker and background intensities can be computed. The values $\mathtt{S}[\check{p}^i]$, where $\check{p}^i \in \Phi_{\mathrm{M}}$, are sorted in a descending order to obtain a sequence $s_1 \geq s_2 \geq \ldots$. Then, the mean intensity of the marker points, mean$(\cdot)$, is computed as:

$$\mu_{\mathrm{M}} = \mathrm{mean}\{s_1, \ldots, s_{n_{\mathrm{M}}}\}. \tag{4.1.5}$$

Finally, the mean and standard deviation values, std$(\cdot)$, of the background points are computed over the intensities of points $\forall \check{p}^j \in \Phi_{\mathrm{B}} \subset \Gamma^2$ as:

$$\mu_{\mathrm{B}} = \mathrm{mean}\{\mathtt{S}[\check{p}^j]\}, \quad \sigma_{\mathrm{B}} = \mathrm{std}\{\mathtt{S}[\check{p}^j]\}. \tag{4.1.6}$$

We select Φ_{M} and Φ_{B} as non-intersecting regions in the image. Further details about the choice of Φ_{M} and Φ_{B} are discussed in Section 4.1.3.

4.1.3 Choice of Temporal Filter

In general, it is expected that the CNR of the stationary components in the filtered images improves with the increase in the number of unfiltered frames that were used to create the filtered image. However, the use of the temporal filtering can, in fact, blur the markers when motion is present, thus degrading the temporal resolution of the filtered image sequence. Also, blurred markers are prone to loosing contrast on their local background, which makes successful image registration more challenging. Consequently, we aim to compare the ability of the filters to enhance image quality when the marker motion is expected. This section presents the results of application of the temporal filters listed in Section 4.1.1 to fluoroscopic sequences of the Quasar phantom with linearly moving fiducial markers.

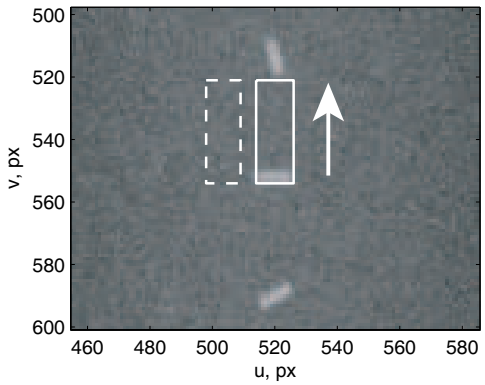


Figure 4.2: The choice of Φ_{M} (*solid line rectangle*) and Φ_{B} (*dashed line rectangle*) for the middle marker depicted over the first frame of the sequence. The *white arrow* shows the direction of the marker motion.

Moving markers data set, and the choice of Φ_{M} and Φ_{B}

The Quasar phantom was described in Section 2.3.2, and its insert in Section 7.1.5. To determine the most appropriate temporal filter, we used a fluoroscopic sequence with the markers moving linearly along the direction of the insert's axis (the V axis in the UV coordinate system), as shown in Figure 4.2.

Each marker ROI, Φ_{M}, was selected as a rectangle that includes one marker, located at its bottom in the first frame of a sequence, given that the motion is performed upwards, and stretches up to the next marker, not including it (see Figure 4.2). The marker points used in CNR computation are selected as n_{M} brightest points in Φ_{M} (see Section 4.1.2). The filtered image is created from the frames $\mathcal{S}_0, \ldots, \mathcal{S}_{\ell-1}$, and Φ_{M} is the same for all frames. Such an ROI selection in this fluoroscopic sequence, combined with the knowledge of the marker speed,[5] allows us to analyze CNR of the filtered images created from up to 17 unfiltered frames.

The initial fluoroscopic sequence contained 40 frames, from which 23 sequences of 17 frames each were generated, i.e., the first generated sequence includes the frames $\mathcal{S}_0, \ldots, \mathcal{S}_{16}$, the second $\mathcal{S}_1, \ldots, \mathcal{S}_{17}$, and so forth. In the following, we will refer to this "moving markers" data set as "MM 23". The marker and background ROIs, Φ_{M} and Φ_{B}, were selected manually in the first generated sequence based on visual examination, and were computed for the following generated sequences based on the marker speed. Both Φ_{M} and Φ_{B} are the rectangles 15×36 pixels in size placed 4 pixels apart.

[5]The velocity of the markers is about 2.3 mm/sec, and with the image acquisition rate of 5.5 frames per second, the markers travel the inter-marker distance of 10 mm in about 24 frames. A shorter sequence of 17 frames has to be used to exclude the neighbouring markers from Φ_{M}.

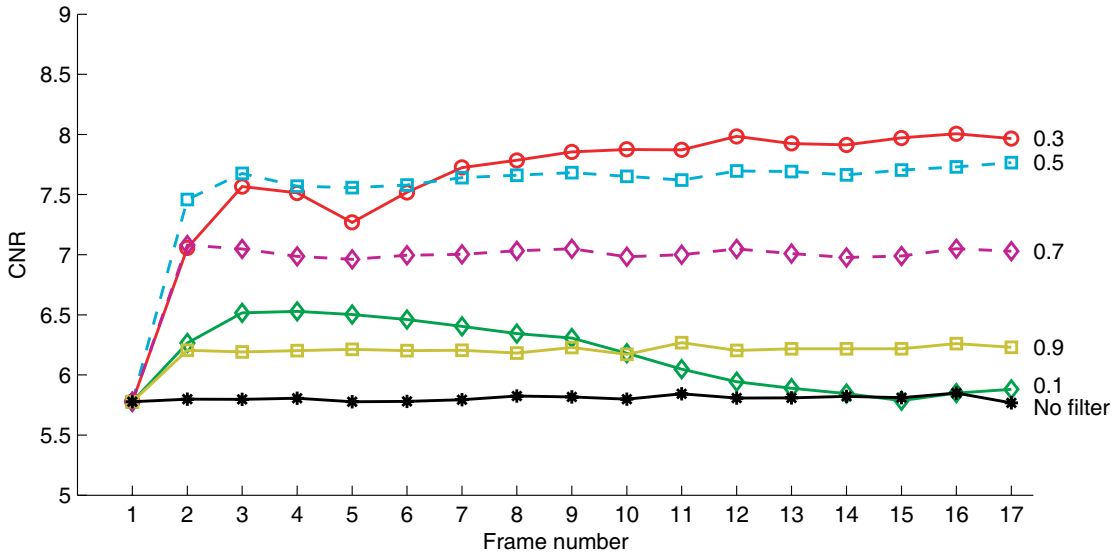


Figure 4.3: CNR in images processed by the temporal recursive filter with $\lambda = 0.1$, 0.3, 0.5, 0.7 and 0.9 (*numbers on the right*). The *"no filter" line* shows the CNR values of the frames before filtering.

Although the phantom contains seven markers, this section reports the CNR results on five: The top and bottom markers had to be removed due to the inclusion of high-contrast features belonging to the phantom insert into Φ_{M} and Φ_{B} in some of the frames, that caused unreasonable values of the CNR.

Recursive filter

Figure 4.3 presents the CNR results after the application of the recursive filter defined in Equation (4.1.1) to the images of the "MM 23" data set. The reported CNRs are the least values among the markers for each of the filtered images, and median values among all sequences. The medians of the CNRs of the unfiltered original frames across all sequences are shown by a *"no filter" line*. For most values of the current frame weight, λ, the recursive filter offers a rapid improvement in the CNR, which saturates after the third frame. For this data set, the CNR was higher for smaller values of λ (*numbers to the right* in Figure 4.3), except for small values of $\lambda < 0.3$.

In addition to the CNR analysis, we also conduct a visual examination of the filtered frames. Figure 4.4 (a) shows a first unfiltered frame of one of the fluoroscopic sequences. Figures (b) and (c) show the 3^{rd} filtered frame with $\lambda = 0.5$ and 13^{th} filtered frame with $\lambda = 0.3$, respectively, both of which correspond to the peak CNR values in that sequence. Visual analysis is important as the CNR itself only shows

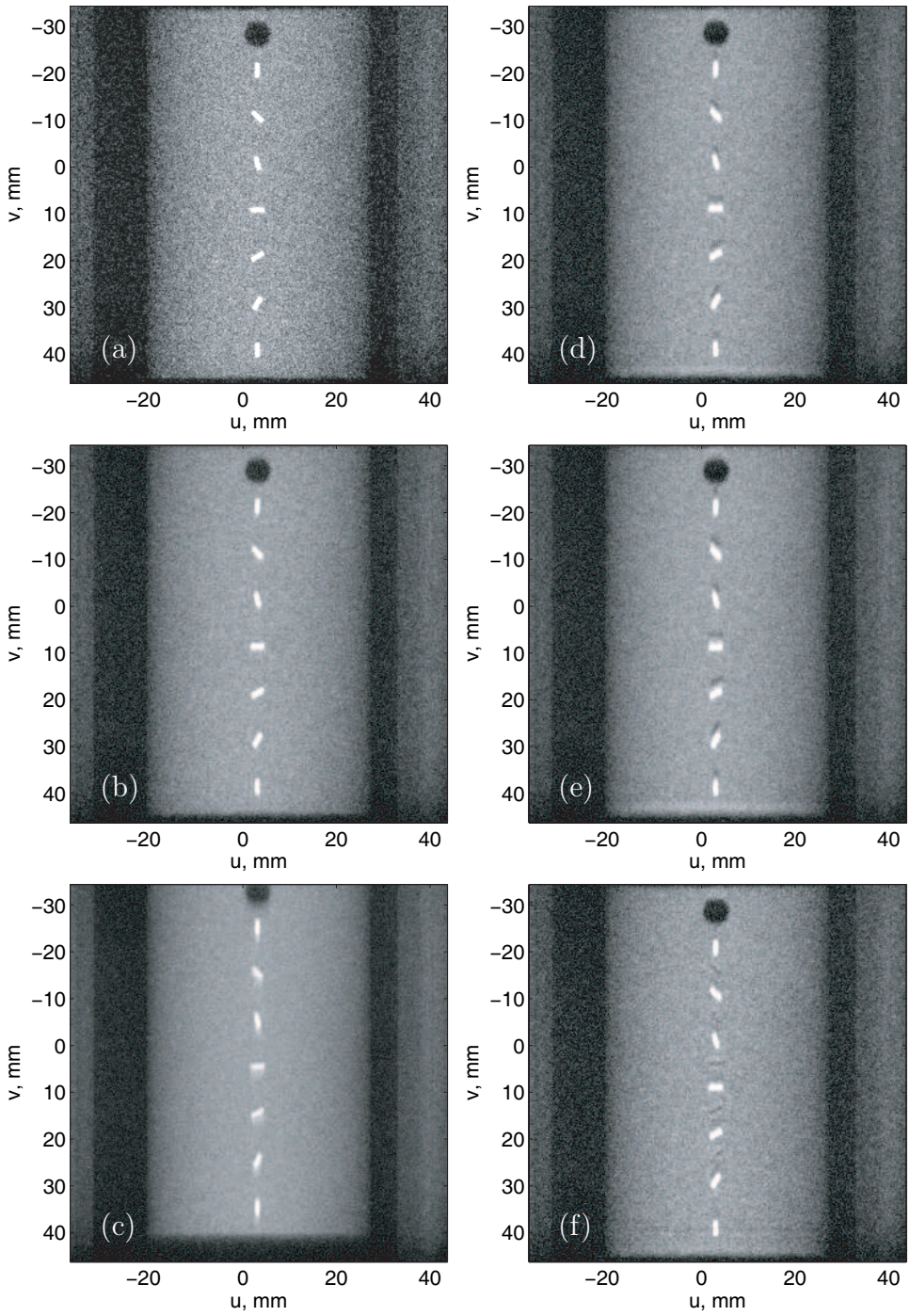


Figure 4.4: The first unfiltered frame of a sample fluoroscopic sequence (a) and the results of filter application: (b) 3^{rd} recursive frame $\bar{\mathcal{S}}_2$ with $\lambda = 0.5$; (c) 13^{th} recursive frame $\bar{\mathcal{S}}_{12}$ with $\lambda = 0.3$; (d) The KLT filter applied to a sequence of $m = 7$ frames with the eigenimage cutoff at $k = 2$ frames; (e) KLT of $m = 9$ with $k = 3$; (f) KLT of $m = 17$ with $k = 6$. The images presented in (b) – (f) are sample filtered images with high CNR values.

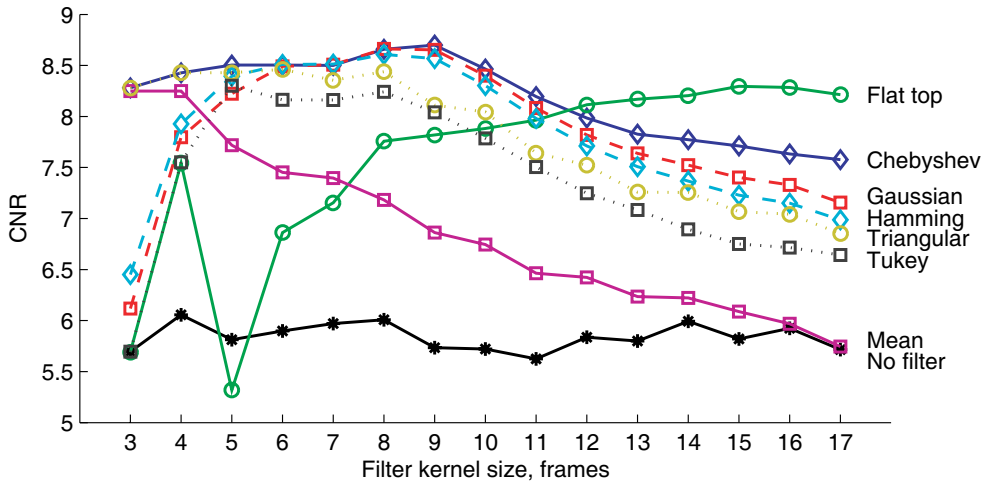


Figure 4.5: CNR in images processed by the finite impulse response (FIR) filters (*titles on the right*). For each kernel size, ℓ, the *"no filter" line* shows the CNR values of the unfiltered frames with the largest coefficient.

the ratio between the contrast and the noise but can not always directly reflect on the blurriness of the markers,[6] which we aim to keep to the minimum, as otherwise the geometric uncertainty and reliability of a subsequent registration may degrade. For example, the comparison between the images in Figures 4.4 (b) and (c), which have very similar CNR values, demonstrates that though markers are undesirably blurred in image (c), a high CNR is still achieved by oversmoothing the background.

FIR filters

This section presents the results of application of various FIR filters to the images of the "MM 23" data set described earlier. We consider the Dolph-Chebyshev, flat top, Gaussian, Hamming, mean, triangular, and Tukey filters, the detailed descriptions of the which are given in Appendix C.2.

Figure 4.5 displays the CNR results of the application of the FIR filters to the images depending on the filter kernel size, ℓ, that was defined in Section 4.1.1. Each line corresponds to the CNR values of one filter, that were computed as the minimal CNR among the markers in each filtered image, and then the median across the images in different sequences. A *"no filter" line* shows the median across the sequences of the minimal CNR across markers that had the largest window coefficient

[6] Blurriness explicitly influences the CNR when the fiducial markers move so fast that they do not overlap in the neighbouring frames, and the markers in the filtered image start to loose their intensity.

for each ℓ, i.e., the middle frame $(\ell+1)/2$ for odd ℓ or the mean between the least CNRs of two middle frames $\lfloor(\ell+1)/2\rfloor$ and $\lceil(\ell+1)/2\rceil$ for even ℓ. The filter kernel size is used in accordance to the notation accepted by Harris [80], where ℓ does not always correspond to an effective kernel size: The definition of some filters is such that $w_0 = w_{\ell-1} = 0$ (e.g., Bartlett-Hann, Bartlett, Bohman, Hann, Tukey) or are very small comparing to the other coefficients (e.g., Blackman-Harris, flat top). This means that at the same ℓ these filters have a potential of a lower CNR comparing to other filters whose side coefficients are larger, and hence, side frames $\mathcal{S}_0$ and $\mathcal{S}_{\ell-1}$ have higher contribution to the filtered image.

We have also considered the following filters defined in [80, 1], which were excluded from presentation of the results due to their like performance to at least one of the seven presented filters:

- The Kaiser filter has essentially the same behaviour as the mean filter, with slightly higher CNR;
- The performance of the Bartlett-Hann, Bartlett, and Hann filters conforms to that of the Gaussian filter with either lower or comparable CNRs;
- The Parzen, Bohman and Blackman filters provide similar CNR values to those of the Gaussian filter for $1 \leq \ell \leq 7$ and to those of the Dolph-Chebyshev filter for $8 \leq \ell \leq 17$;
- The Taylor filter produces the CNR values similar to those of the triangular filter for $1 \leq \ell \leq 7$ and to those of the Tukey filter for $8 \leq \ell \leq 17$.

From Figure 4.5, we conclude that, among the considered FIR filters, the Dolph-Chebyshev window is the best choice for the fluoroscopic sequences for which the marker motion is expected as it quickly improves the CNR even for small values of $\ell < 5$. The application of this filter keeps the CNR higher or comparable to other filters for larger values of ℓ, with the exception of the flat top filter at $\ell \geq 12$. In Figures 4.6 (a), (b), and (c) we present the results of the Dolph-Chebyshev filter application at $\ell = 3$, 5, and 9, respectively, to one of the fluoroscopic sequences in the "MM 23" data set. For comparison, we also show the results of the mean filter application to the same sequence in Figures 4.6 (d) – (e), as it was reported to be used in temporal filtering of fluoroscopic sequences [7]. It can be seen that lower CNRs in the images processed by the mean filter are caused by considerable and undesirable blurring of the markers that decreases the intensities of the marker points.

The KLT method

One of the conceptual differences between the KLT and recursive or FIR filtering is that the KLT method produces a filtered sequence of m frames (simultaneously)

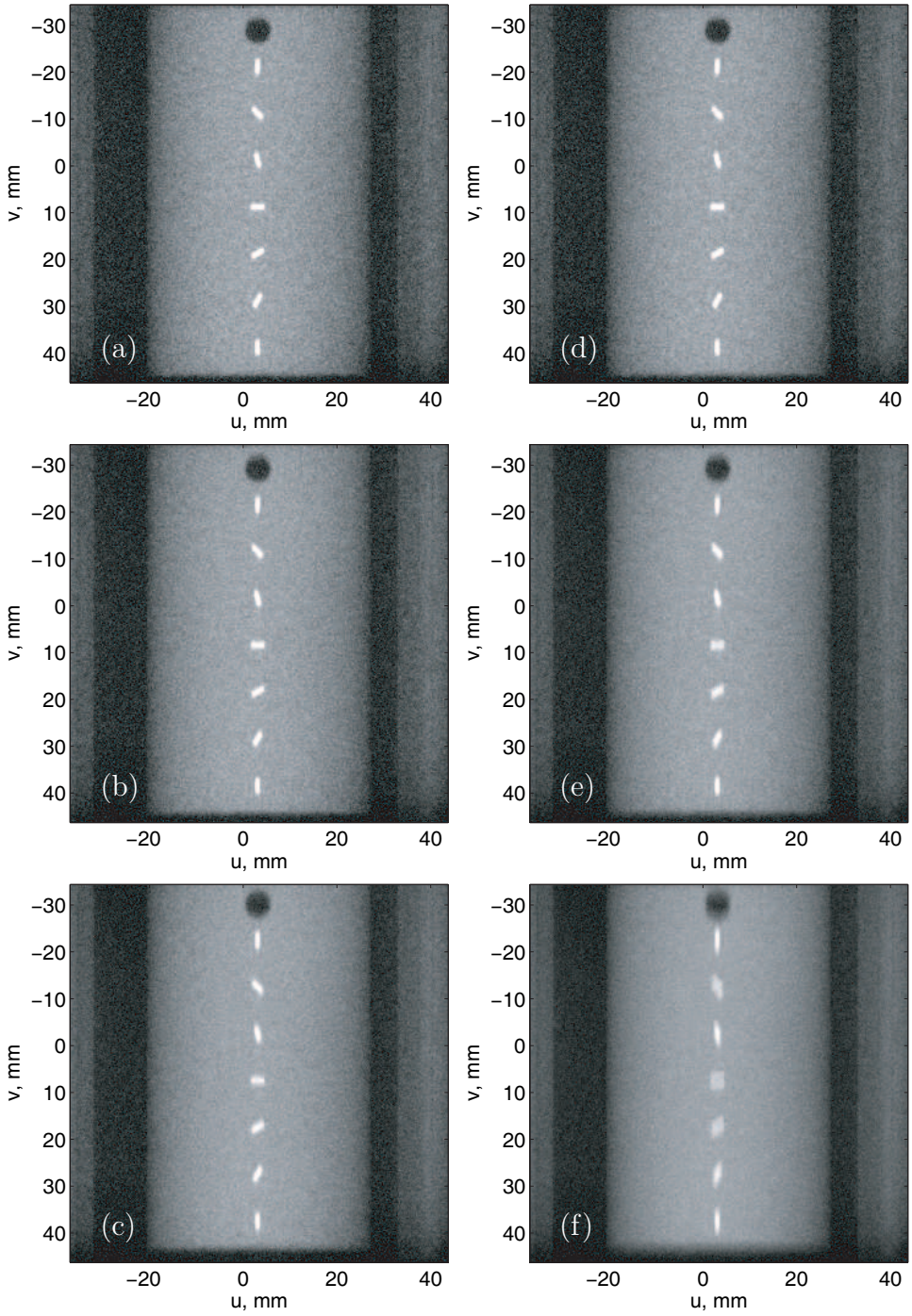


Figure 4.6: The results of filter application: Dolph-Chebyshev filter with kernel sizes $\ell = 3$ (a), 5 (b), and 9 (c) and the mean filter with $\ell = 3$ (d), 5 (e), and 9 (f). The unfiltered first frame of the sequence is shown in Figure 4.4 (a).

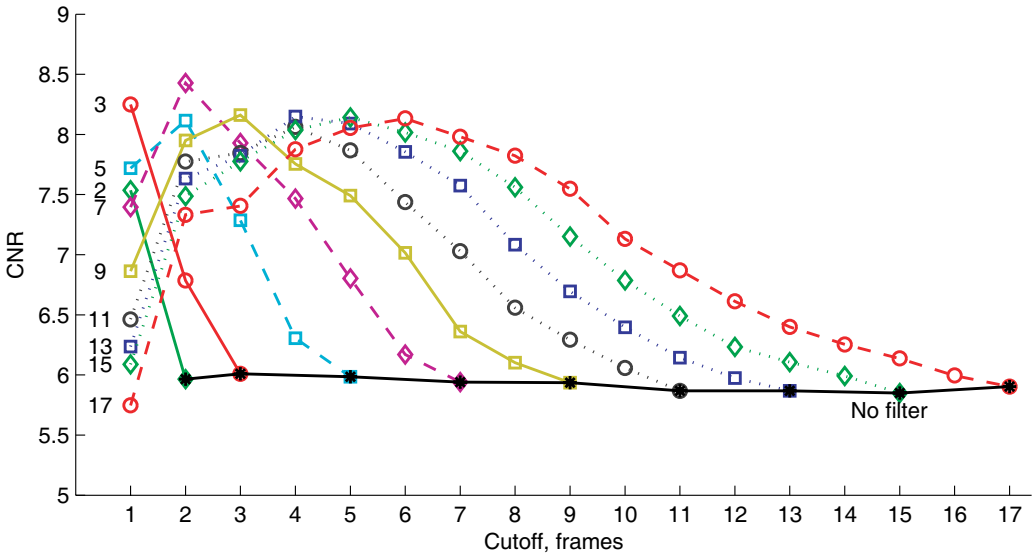


Figure 4.7: CNR in images processed by the KLT filter. The *numbers on the left* show the length of the unfiltered image sequence, m, and the *horizontal axis* corresponds to the cutoff, k. The *"no filter" line* shows the median CNR values of unfiltered frames, where $m = k$.

from an unfiltered sequence of the same length. This means, that, contrary to any FIR filter, where for each ℓ only one filtered image was produced from each of the 23 sequences, the KLT method produces m filtered images for every m and k value, where k is a *cutoff*, or the number of eigenimages left after the filtering, as described in Section C.1. Hence, to report the results, we compute the median CNRs across m filtered images for each marker, then the minimal CNR across the markers, and finally, the median CNR across the sequences. Figure 4.7 presents the resulting CNR values. When $m = k$, no filtering is performed, which is shown by a *"no filter" line*. The *horizontal axis* shows the cutoff k, and the *numbers on the left* of the lines correspond to m.

The analysis of the sample filtered frames in Figures 4.4 (d) – (f), that attain peak CNR values in their respective image sequence, reveals the trend to introduce shadows of the markers from the neighbouring frames. This results in an incorporation of artificial minima and maxima in the image intensity function. This phenomenon is quite well known in image compression techniques that are based on related ideas, and is known as "ringing", e.g., [209]. Since we are seeking to filter the images in order to facilitate accurate registration, this ringing behaviour is rather undesirable.

Summary

The recursive, FIR, and KLT filtering methods described above have the following advantages and drawbacks:

- **Recursive:** The CNR depends on the current frame weight, λ, which should be selected according to the marker speed. This means that should the markers move at varying speeds, which is likely to happen in patient image sequences, the CNR of the markers will change rapidly, as well.
- **FIR:** As explained above, we concluded that a sensible choice among different FIR filters is the Dolph-Chebyshev filter due to its ability for substantial CNR improvement even for small values of ℓ, and competitive CNRs for larger values of ℓ. Even in the presence of motion, the Dolph-Chebyshev filter tends to offer high CNR values for a wide range of ℓ values.
- **KLT:** This method tends to be more expensive computationally than the FIR or recursive filtering due to the necessity of eigenvalue computation. It is prone to introducing the artificial shadows of the markers (ringing) that may later mislead the registration. For fair comparison, it should be mentioned that both FIR and recursive filters can also produce images with multiple markers or shadows in lieu of one. However, that happens for faster moving markers when their measurements in the consecutive frames do not overlap.

As can be seen from the results of applying different filters to the images of the "MM 23" data set presented above, the Dolph-Chebyshev filter outperforms both the recursive and KLT methods in terms of the CNR and visual assessment. It produces higher CNRs for a wider range of ℓ values, making it better suited for fluoroscopic sequences where marker motion speed is not known a priori. Hence, we use the Dolph-Chebyshev temporal filter to preprocess the fluoroscopic images in the proposed marker localization and tracking procedures. The selection of the filter kernel size, ℓ, is discussed in the context of marker localization success rates in Section 5.4.

4.2 Marker Amplification using Existing Techniques

As explained in Section 1.1.5, the positions of the fiducial markers represent an accurate estimate of the target position. Therefore, it is essential that image registration in both marker localization and tracking procedures aligns the corresponding markers in the images, even if other features are not ideally aligned. This requirement leads to the necessity of a method for amplification of the markers and suppression of the background in fluoroscopic images. Given the local nature of the marker contrast (see

an example in Figure 4.8 (d)), we conclude that there is a need for a digital filter that will conduct a local intensity analysis to amplify the features that:

- possess local contrast by having intensity values that are significantly higher than their immediate background, where significance is assessed relative to the noise variation;
- are characterized by a certain known size and shape.

This section presents an overview of the existing feature enhancement methods.

We start with the discussion on *morphological opening* [98] and *template matching* [78, 43, 169, 31, 7, 174, 141] in Section 4.2.1. We also consider the *marker extraction kernel (MEK)* [160, 162] that was developed for detection of markers in portal images and adapted or used directly in other studies [204, 159, 218, 217, 108]. Then, Section 4.2.2 outlines the scale-space framework [127], which has found wide utilization in the enhancement of vessels and other features [63, 189, 121, 132], and in feature detection and image matching applications for computer vision [134, 126, 127, 47, 170, 35, 96, 148, 215].

Although we usually validate algorithms on phantom data sets, patient images are used in this and the following sections to review existing strategies and design the marker enhancement filter. We have done this because the characteristics of real patient images are crucial for this step (the phantom images often possess a better contrast than the patient images).

4.2.1 Existing Feature Enhancement Approaches

Morphological opening

The *morphological opening* technique [73] aims to extract features of a particular size and shape in the image by using a structuring element $\mathcal{E}$ represented by a binary image, where 1's define the element of some predefined shape and 0's the background. Often, it is used for improving the signal in noisy binary images. Essentially, morphological opening removes (by flattening into the background) measurements smaller than $\mathcal{E}$ by successively applying the following operations to each pixel of the input image $\bar{\mathcal{S}}$:

1. **Erosion**: The value of the output pixel is the *minimum* value of all the pixels in the input pixel's neighbourhood, defined by the 1's in $\mathcal{E}$;

2. **Dilation**: The value of the output pixel is the *maximum* value of all the pixels in the input pixel's neighbourhood.

While erosion removes small measurements and the noise, dilation restores the larger measurements to their original size.

To process the images with the measurements of the cylindrical fiducial markers, we have used a rectangular structuring element. To address the issue of marker orientation, multiple structuring elements have to be generated, $\mathcal{E}_{\phi_i}$, where ϕ_i is the orientation angle. The application of the morphological opening with multiple elements creates a number of images:

$$\mathcal{S}_{\phi_i} = \text{imopen}(\bar{\mathcal{S}}, \mathcal{E}_{\phi_i}), \tag{4.2.7}$$

which we can combine by applying the maximum intensity projection (MIP) [56] operation

$$\mathcal{Y} = \text{MIP}\{\mathcal{S}_{\phi_i}\}, \tag{4.2.8}$$

where each point p in the resulting image $\mathcal{Y}$ is computed as $\mathcal{Y}(p) = \max_i\{\mathcal{S}_{\phi_i}(p)\}$.

Figure 4.8 (b) demonstrates the MIP result of the application of the morphological opening (implemented using the Matlab function `imopen`) to a sample fluoroscopic image of a prostate patient shown in Figure 4.8 (a) that was preprocessed by the Dolph-Chebyshev temporal filter with kernel size $\ell = 7$. We used a rectangular structuring element 12×3 pixels in size, which is similar to the marker dimensions. The structuring elements corresponding to different orientation angles, $\mathcal{E}_{\phi_i}$, were generated from the initial rectangle by rotating it around its centre, and using the nearest neighbour interpolation (recall that the structuring element is a binary image). The rotation angle was discretized at a 5° step, i.e., $\phi_i = -85^\circ, -80^\circ, \ldots, 90^\circ$.

Although the application of morphological opening to the image results in a significant noise reduction, as can be seen in Figure 4.8 (b) and its corresponding intensity profile in (e), the markers possess only local contrast, while large intensity variations of the background and anatomical features are preserved. Therefore, the morphological opening alone is insufficient for our purposes. Other choices for the structuring element, such as a circle, produced very similar results and hence are not included in this discussion.

Template matching

Template matching is a technique for finding the points of maximal correspondence between an image under consideration and a template image, which is often smaller in size and restricted to a measurement of the object that is being sought in the larger image. The template is compared to the parts of the image, and the goodness of the match is usually quantified by the means of normalized correlation. The template can be extracted from a similar image [43, 169, 174], synthesized based on the known physical properties of the object and intensity characteristics of a typical object's measurement [78, 31], or by using other available images of the patient [7]. The result

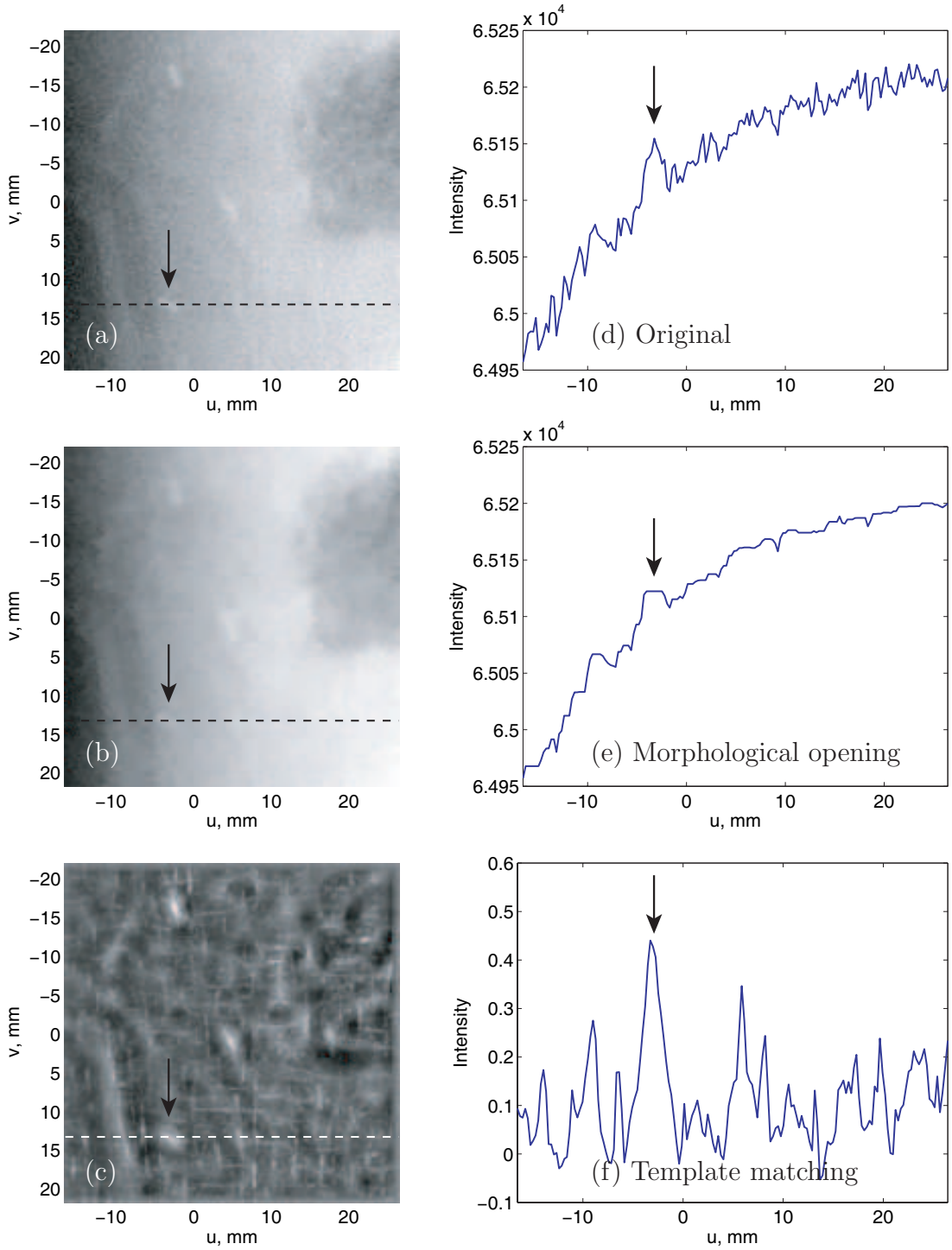


Figure 4.8: The results of the application of morphological opening (b) and template matching (c) to the prostate patient fluoroscopic image preprocessed by the temporal Dolph-Chebyshev filter with $\ell = 7$ kernel size (a). Figures (d), (e), and (f) show the corresponding intensity profiles at rows in (a), (b), and (c), respectively, marked by the *dashed lines*. *Arrows* point to one of the three markers.

of template matching can be interpreted as an image that has high intensities in the points where the image contains features similar to the template. As such, this technique can be used for feature enhancement.

Figure 4.8 (c) shows an image that was produced by computing the normalized cross-correlation between the templates and a sample fluoroscopic image of a prostate patient shown in (a). The multiple templates $\mathcal{E}_{\phi_i}$ were the same that were used for the morphological opening described above, with the same discretization step and combination of $\mathcal{S}_{\phi_i}$ images by the MIP operation defined in Equation (4.2.8). The analysis of the resulting image and its intensity profile in Figure 4.8 (f) shows that template matching helped to suppress low-frequency background intensity variation, and the markers now possess a better contrast. However, the difference in the intensity between the markers and other features is still very small, and the noise is high, which may result in unreliable detection and tracking, making a conventional template matching approach unsuitable for our purposes.

Marker Extraction Kernel (MEK)

The MEK was designed by Nederveen *et al.* [160] to enhance the measurements of the gold cylindrical fiducial markers of 5×1.2 and 5×1.4 mm in the portal images. Its parameters are adjusted so that it gives zero response on a constant intensity background and reaches its maximum when calculated in the centre of the higher intensity feature that is similar in size, shape, and orientation to the marker. When applied to an image, the MEK quantifies how likely the feature is to be a marker by taking into consideration the average intensities I_{ab}, I_{ϵ}, and I_{δ} inside the central part of the feature, L_{ab}, and its border areas, L_{ϵ} and L_{δ}, respectively, shown in Figure 4.9. The MEK response, referred to as the marker value, MV, is calculated as [160]:

$$\mathrm{MV} = \frac{w}{1+w} I_{ab} + \frac{1}{1+w} I_{\epsilon} - I_{\delta}, \tag{4.2.9}$$

where w is a weight parameter. The parameters a and b are selected to approximate the width and length of the marker in the projection plane. To calculate the MV for a particular pixel $\check{p} \in \Gamma^2$, the kernel's centre is positioned at $\check{p}$, and the MEK responses $\mathrm{MV}_{\phi_i}(\check{p})$ are computed for different orientations of the kernel described by the rotation angles ϕ_i. The pixels from the neighbourhood of the point $\check{p}$ belong to L_{ab}, L_{ϵ} or L_{δ} if their centres (grid points) belong to these areas. Finally, all $\mathrm{MV}_{\phi_i}(\check{p})$ are combined by either computing their maximum, as described in the original method [160, 162], or sometimes their sum [159].

The MEK was used for marker enhancement and detection in a number of studies, both on its own [204, 218, 217, 108], and in combination with other methods,

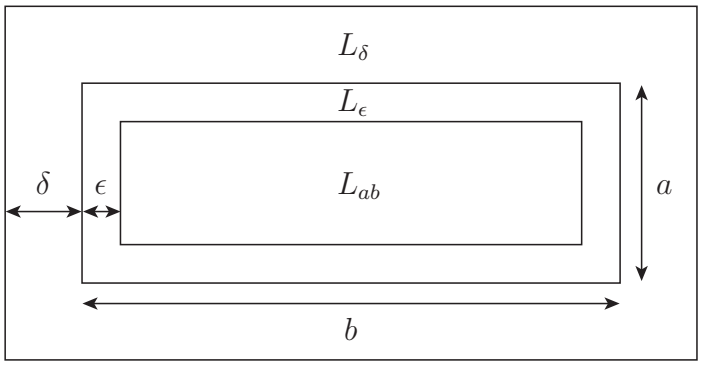


Figure 4.9: Marker structure exploited by the marker extraction kernel (MEK). The filter response is calculated using the average intensity values in the central part of the measurement, L_{ab}, and its inner and outer border areas, L_ϵ and L_δ with border sizes ϵ and δ, respectively. The parameters a and b relate to the expected marker size. This schematic is adapted from Nederveen *et al.* [160].

such as template matching [159, 97]. To reduce the expense of calculating the MV for multiple orientations, knowledge about the marker orientation, that is obtained from the portal images acquired during the first treatment fraction of a patient, can be used [204]. Alternatively, the MVs can be computed only over a small range around some preliminary estimated orientations [160, 162] that are derived from the orientation of the phantom. As suggested by Nederveen *et al.* [160, 162], in the clinical application the orientations can be found from the CT images and marker geometry.

In our experiment, we assumed that marker orientations were not known a priori. The rotation angle was discretized at a 5° step, i.e., $\phi_i = -85°, -80°, \ldots, 90°$. The parameters assume the following values: $a = 0.8$ mm, $b = 3$ mm, and, as suggested in [160], $w = 3$, $\delta = 2\delta_{\text{px}}$, $\epsilon = \delta_{\text{px}}$, where δ_{px} is the pixel size. Although the size of the markers at the isocentric plane can be slightly different from the fiducial marker size, it will be reasonably similar, as the fiducial markers are positioned in the close proximity to the LINAC's isocentre. Also, as verified in [160] and confirmed by our experiments, the variation of a and b over a few pixels is not critical to the MEK response.

Figure 4.10 presents the results of the MEK application to the fluoroscopic image of a prostate patient shown in Figure 4.8 (a). It is rather clear that in fluoroscopic images, the MEK is more successful in marker enhancement than the morphological opening and template matching described above. The MEK flattens the background, and has significantly stronger response to the marker-like measurements than to anatomical features. Therefore, a global threshold can be employed after MEK application to separate the markers from the background and other features.

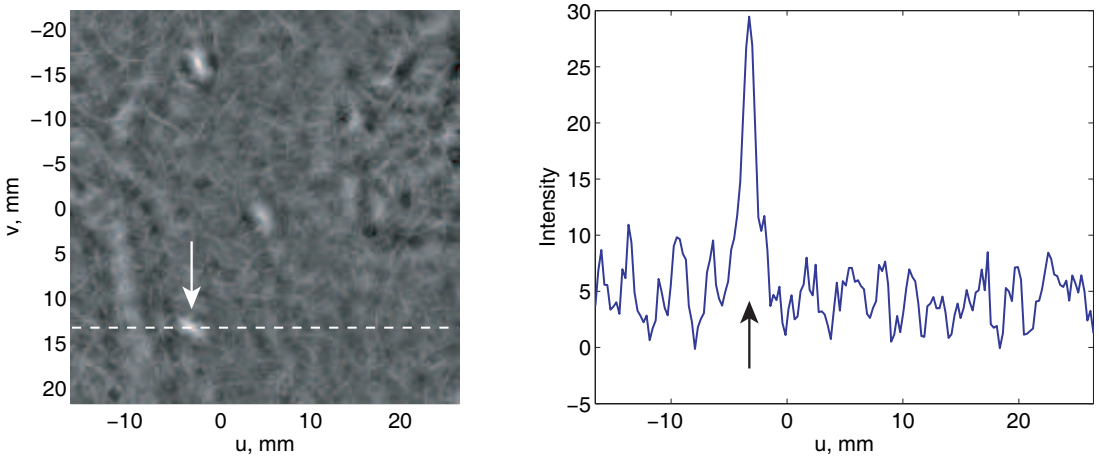


Figure 4.10: The result of the MEK application (*left*) to the image of the prostate patient displayed in Figure 4.8 (a), and its corresponding intensity profile (*right*).

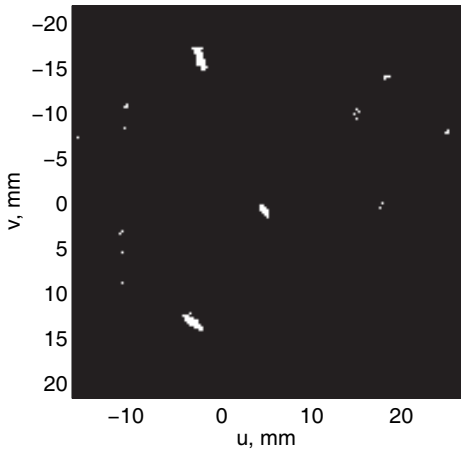


Figure 4.11: Global thresholding of the image processed by the MEK (Figure 4.10, *left*).

Figure 4.11 demonstrates the result of the global thresholding of the image shown in Figure 4.10 (*left*), which was performed as follows: The intensity value of 1 was assigned to n_{M} brightest points in the image, and 0 to the rest. The value n_{M} is an estimate of the number of pixels that constitute the markers based on knowledge of the marker size and number, and is computed as defined in Equation (4.1.4) with $\kappa = 1$. Although the markers are greatly enhanced, there is a number of false positives in the image. Also, given that the marker orientation is not known, the MEK computed from multiple orientations distorts the shapes of the markers, which can potentially decrease the accuracy of the subsequent image registration. These reasons led us to the exploration of another approach, namely, the scale-space analysis described in the next section.

4.2.2 Scale-Space Representation and Analysis

In the fields of computer vision and image understanding, there is a broadly accepted perspective that features in the images exist as meaningful entities only over a limited range of scales [106, 127]. Hence, to perform an automatic image interpretation, such as segmentation and feature detection for image matching, a *multi-scale representation* was proposed [235, 106]. In our case, the fiducial markers are a part of the standard clinical procedure, and hence their size is known in advance. Therefore, we use general principles of the multi-scale analysis explained below, but we produce images at a single preselected scale.

In the scale-space framework, the original continuous image $\mathcal{A} : \mathbb{R}^d \to \mathbb{R}$ is embedded into a one-parameter family of derived images $\mathcal{L} : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}$, where $\sigma \in \mathbb{R}_+$ is a *scale parameter*.[7] The *scale-space representation* $\mathcal{L}$ for $\forall p \in \mathbb{R}^d$ is defined as:

$$\mathcal{L}(p, \sigma) = \begin{cases} \mathcal{A}(p) & \text{for } \sigma = 0, \\ \mathcal{A} * \mathcal{G}(p, \sigma) & \text{for } \sigma > 0, \end{cases} \tag{4.2.10}$$

where the Gaussian is defined as:

$$\mathcal{G}(p, \sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{\|p\|^2}{2\sigma^2}\right). \tag{4.2.11}$$

That is, representation for $\sigma = 0$ is the original image itself and for $\sigma > 0$ it is computed as the convolution with the Gaussian of the corresponding scale. The convolution with the Gaussian generates a scale-space representation characterized with the following property: New features should not be created when σ is increased [106] (i.e., when the resolution is diminished, or the image is blurred).

[7] $\mathbb{R}_+$ is a set of nonnegative real numbers.

A common approach to analyze a local behaviour of an image is to consider its second-order Taylor expansion in the neighbourhood of a point p:

$$\mathcal{L}(p+\delta p,\sigma) \approx \mathcal{L}(p,\sigma) + \delta p^T \nabla\mathcal{L}(p,\sigma) + \frac{1}{2}\,\delta p^T H(p,\sigma)\delta p, \tag{4.2.12}$$

where $\nabla\mathcal{L}(p,\sigma)$ is the gradient vector, and $H(p,\sigma)$ is the Hessian matrix of $\mathcal{L}$ computed at p at a scale σ. The first and second partial derivatives of $\mathcal{L}$ for the construction of the gradient, $\nabla\mathcal{L}(p,\sigma)_i = \frac{\partial}{\partial p_i}\mathcal{L}(p,\sigma)$, and the Hessian, $H(p,\sigma)_{ij} = \frac{\partial^2}{\partial p_i \partial p_j}\mathcal{L}(p,\sigma)$, $i,j = \{1,\ldots,d\}$, can be computed using convolution with the Gaussian derivatives [127]:

$$\frac{\partial}{\partial p_i}\mathcal{L}(p,\sigma) = \mathcal{A} * \frac{\partial}{\partial p_i}\mathcal{G}(p,\sigma) \tag{4.2.13}$$

and

$$\frac{\partial^2}{\partial p_i \partial p_j}\mathcal{L}(p,\sigma) = \mathcal{A} * \frac{\partial^2}{\partial p_i \partial p_j}\mathcal{G}(p,\sigma). \tag{4.2.14}$$

To simplify the notation, in the following we will use $\mathcal{L}_u$, $\mathcal{L}_v$, $\mathcal{L}_{uu}$, $\mathcal{L}_{uv}$, and $\mathcal{L}_{vv}$ to denote first and second partial derivatives in the U and V directions of the 2D UV coordinate system, namely, $\frac{\partial}{\partial u}\mathcal{L}(p,\sigma)$, $\frac{\partial}{\partial v}\mathcal{L}(p,\sigma)$, $\frac{\partial^2}{\partial u^2}\mathcal{L}(p,\sigma)$, $\frac{\partial^2}{\partial u \partial v}\mathcal{L}(p,\sigma)$, and $\frac{\partial^2}{\partial v^2}\mathcal{L}(p,\sigma)$, respectively.

Due to the fact that the fiducial markers are all of equal or very similar size to each other, we conclude that it suffices to select one fixed scale for marker enhancement or detection based on the scale-space framework. Therefore, we drop the notion of the scale-normalized derivatives[8] used to balance detector responses at different scales, and use the simple derivatives defined in Equations (4.2.13) and (4.2.14) instead.

▲ σ, **Gaussian scale**

The Gaussian scale of $\sigma = 0.7$ mm was selected empirically based on our experiments with the phantom and patient data sets as the one that provides the best preservation of the marker size and shape. For example, compare the results of application of one of the filters discussed later in this section, LoG^2, to a prostate patient image with $\sigma = 0.7$ mm (Figure 4.12 (c)), $\sigma = 0.4$ mm, and $\sigma = 1$ mm (Figure 4.13 (a) and (b)). Smaller scales do not provide enough noise suppression, and create a chopped marker, which is not desirable for a subsequent 2D/3D image registration. On the other hand, as σ increases, the markers become smoothed out and rounded, and larger anatomical features are amplified. ■

In the remainder of this section, we review a number of existing methods that are based on the scale-space approach, and evaluate them in the application to marker

[8] The normalized derivative can be computed as $\frac{\partial}{\partial p_i}\mathcal{L}(p,\sigma)_{\text{norm}} = \sigma^\gamma \mathcal{A} * \frac{\partial}{\partial p_i}\mathcal{G}(p,\sigma)$, where $\gamma = 1$ is a common choice [127].

enhancement in the prostate patient images. The approaches can be split into two groups: blob detection and line-like feature enhancement.

Blob detection in computer vision

The scale-space approach is extensively used in computer vision for the detection of local features. A *local feature* can be defined as an image pattern that is different from its immediate neighbourhood in intensity, color or texture [215]. Local features can include points, edges, and small image patches. Feature identification has found its application in the detection of measurements of specific objects in the images, matching and tracking applications, camera calibration and 3D reconstruction, and ability to recognize objects or scenes without the need for explicit segmentation, among others [215].

A *blob* is a particular type of a local feature that can be usable in marker enhancement or detection. It can be loosely defined as a region that is brighter or darker than its background [125]. In other words, the blob is a region of connected points associated with (at least) one local extremum, where all the points of the region possess intensity higher or lower than their immediate background. The following discussion deals with the bright blobs on a dark background.

While it is usual for feature detectors to only extract the extremum points of the blobs at different scales, we seek to develop a method that will have the whole image as its output rather than a set of points. The sought-for method should provide improvement in the marker contrast, preserve its size and shape, and suppress background and other features in the image.

We considered the following approaches:

- **Laplacian blobs.** Detectors based on the Laplacian of Gaussian are widely used in computer vision. The Laplacian of Gaussian for a 2D image is the trace of the Hessian matrices at each image point p at a particular scale σ. That is,

$$\mathrm{LoG} = \mathcal{L}_{uu} + \mathcal{L}_{vv}, \tag{4.2.15}$$

  where $\mathcal{L}_{uu} = \frac{\partial^2}{\partial u^2}\mathcal{L}(p,\sigma)$ and $\mathcal{L}_{vv} = \frac{\partial^2}{\partial v^2}\mathcal{L}(p,\sigma)$. Blob detectors based on the LoG seek the points of local maxima in $-$LoG [47], |LoG| [126] or LoG^2 [127].

  Filters based on the Laplacian of Gaussians were used to enhance 2D markers in portal images, as reported in several studies [170, 35]. In order to enhance the measurements of gold cylindrical fiducial markers of size 3×1 mm, Park *et al.* [170] convolved the image with a Gaussian defined as $\exp(-\|p\|^2/(2\sigma^2))/\sqrt{2\pi\sigma^2}$, and then the LoG was computed. After this, the intensity threshold was applied to select candidate markers. Finally, to eliminate

false positives and false negatives, the resulting image was registered to the CT image. Buck *et al.* [35] enhanced the measurements of the spherical tungsten markers of 2 and 1.5 mm in diameter by convolving the images with the Mexican hat filter, which was defined as LoG computed for the images convolved with $\exp(-\|p\|^2/(2\sigma^2))$. Then, the markers were found by applying an intensity threshold, and utilizing knowledge about marker size and shape.

- **Hessian blobs.** Detectors based on Hessian blobs seek the local maxima of the determinants of the Hessians:

$$\mathrm{detH} = \mathcal{L}_{uu}\mathcal{L}_{vv} - \mathcal{L}_{uv}^2 \tag{4.2.16}$$

or detH^2 [127, 96].

- **Combined Hessian-Laplace points** are the points where both detH and LoG simultaneously assume local maxima [134, 148, 215]. This is done to reduce the strong response of the LoG-based detectors to edges, as points on the edges are usually characterized with poor localization stability [134]. In [96], the points are defined as:

$$\begin{cases} \operatorname{local}\max_{u,v}\{\mathrm{detH}^2\}, \\ \operatorname{local}\max_{\sigma}\{\mathrm{LoG}\}. \end{cases} \tag{4.2.17}$$

As we are interested in generating the whole image rather than a set of points, we suggest that a filter that will provide significant response at the Hessian-Laplace points can be described as the point-wise multiplication:

$$\mathrm{Cmb} = \mathrm{detH}^2\,\mathrm{LoG} \tag{4.2.18}$$

The images in Figures 4.12 and 4.14 were obtained by applying variations of the Laplacian, Hessian, and combined Hessian-Laplace filters defined above to an image from the patient testing data set. Among them, the LoG^2 method (Figure 4.12 (c)) appears to be the most promising as it efficiently combines the suppression of the noise and the background, including anatomical features, with a reasonable preservation of the marker size and shape. Together with the MEK, we will use the performance of the LoG^2 approach as a benchmark for the proposed filter in the experiments described in Section 7.2.4.

Blob and line-like feature enhancement

Scale-space methods have found wide application in the enhancement of images of blood vessels and bronchi [63, 121]. Retinal images are one prominent application. Other applications include the detection of dendrites, fingerprints, and catheters.

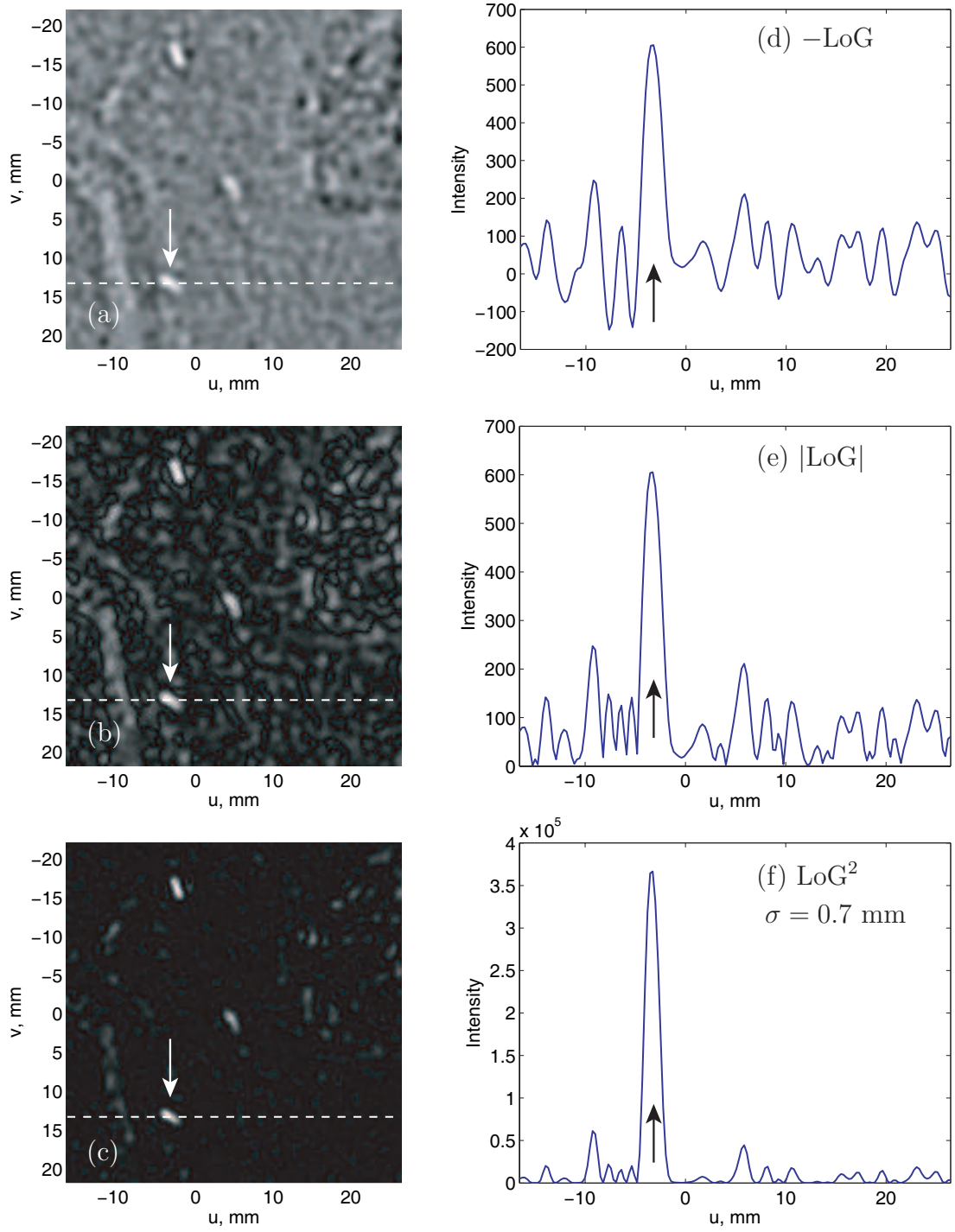


Figure 4.12: The results of the Laplacian blob enhancement −LoG (a), |LoG| (b), and LoG^2 (c) in the image of the prostate patient displayed in Figure 4.8 (a), and their corresponding intensity profiles (d), (e), and (f).

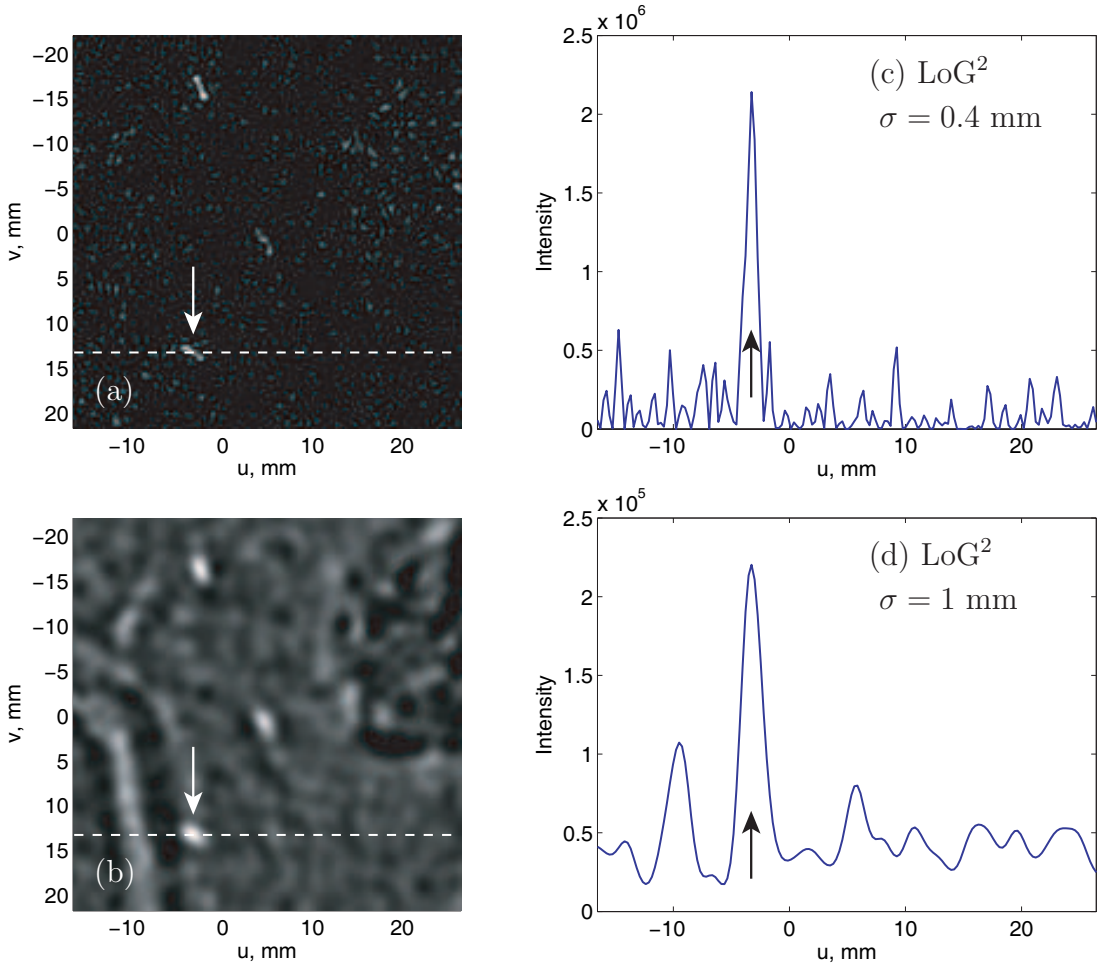


Figure 4.13: The results of the Laplacian blob enhancement LoG^2 with the Gaussian scale parameter $\sigma = 0.4$ mm (a) and $\sigma = 1$ mm (b), and their corresponding intensity profiles (c) and (d).

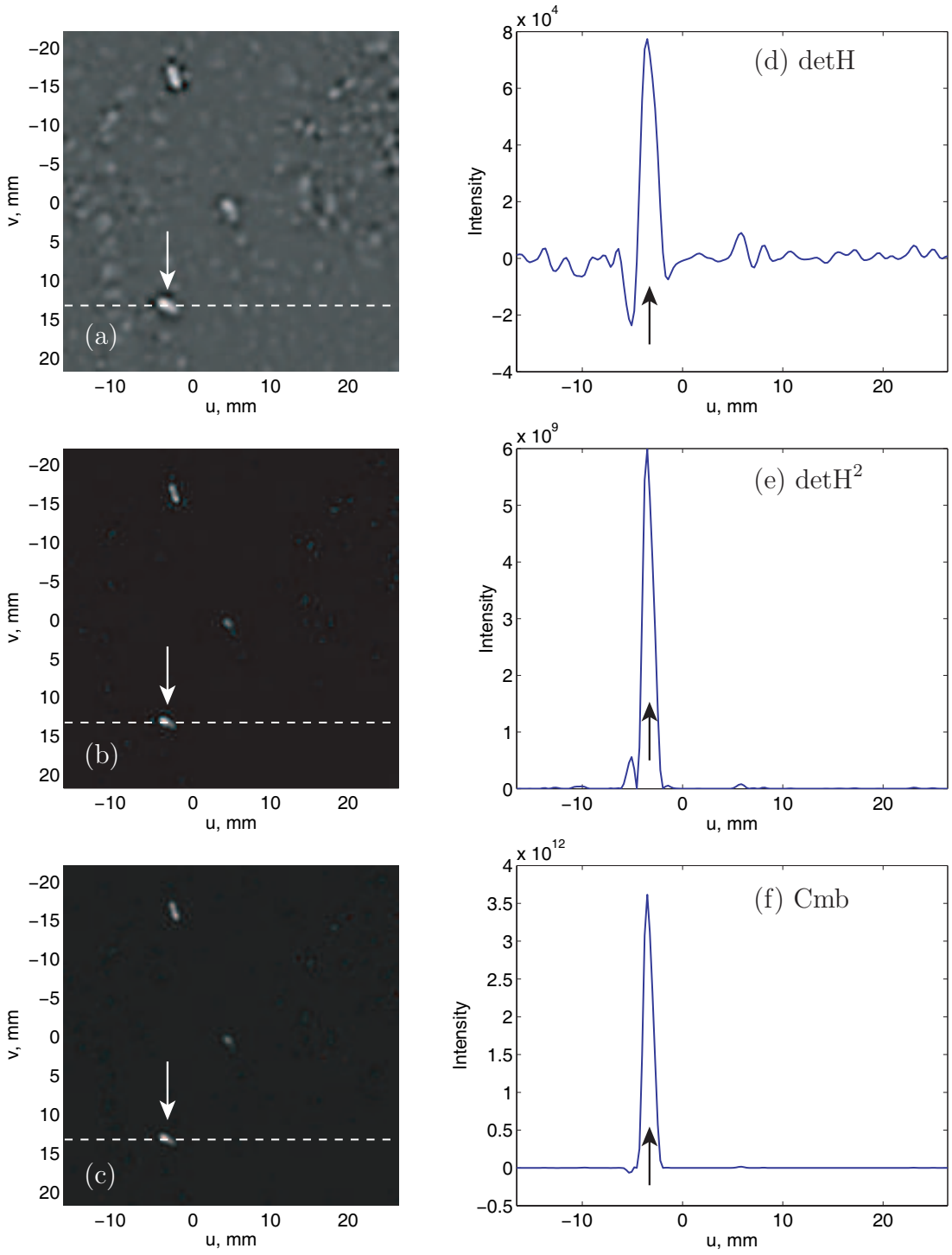


Figure 4.14: The results of the Hessian detH (a) and detH^2 (b), and the combined Hessian-Laplace Cmb (c) blob enhancement in the image of the prostate patient displayed in Figure 4.8 (a), and their corresponding intensity profiles (d), (e), and (f).

The key observation is that the magnitude and ratio of the eigenvalues of the Hessian $H(p,\sigma)$, $\lambda_1 = \lambda_1(p,\sigma)$ and $\lambda_2 = \lambda_2(p,\sigma)$, can characterize the shape of intensity variations in the image. For example, large negative eigenvalues, $\lambda_1 \ll 0$ and $\lambda_2 \ll 0$, indicate a concave shape in the intensity function, i.e., a bright spot on a dark background. The values $\lambda_1 \ll 0$ and $\lambda_2 \approx 0$ of any sign mean that point p belongs to a measurement in the shape of a bright line on a dark background. Different shapes of intensity functions and characteristics of their corresponding eigenvalues are described by Frangi *et al.* [63].

Several feature-enhancement filters that use eigenvalues can be found in the literature, e.g., [132, 189, 63, 121]. To the best of our knowledge, no eigenvalue-based methods that specifically deal with the enhancement of the markers have previously been reported. Most of the known methods are designed for lines and blobs. Cylindrical fiducial markers can be treated as line-like structures that are, unlike vessels, are very limited in length. On the other hand, the marker length and width are not very different, which can make blob-specific filters suited for marker enhancement. In the remainder of this section, we review the methods for line and blob enhancement and compare the results of their application to our patient images. We use the name of the first author of the publication as the name of the corresponding filter for easy identification. The following methods were considered:

- **Lorenz line filter.** This filter is intended for amplification of line-like structures to improve automatic segmentation. Assuming the eigenvalues of the Hessian, $|\lambda_1| \leq |\lambda_2|$, the filter is defined as [132]:

$$\mathcal{Y}_{\text{line}}^{\text{Lorenz}} = \begin{cases} |\lambda_2|, & \text{if } \lambda_2 < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2.19}$$

- **Sato's filter for nodules and vessels.** For this filter, no pre-sorting of the eigenvalues is assumed. We formulate 2D versions of the 3D blob (nodules) and line (vessels) filters presented in [189] in the following way:

$$\mathcal{Y}_{\text{blob}}^{\text{Sato}} = \begin{cases} |\lambda_2| \left(\frac{\lambda_1}{\lambda_2}\right)^{\gamma}, & \text{if } \lambda_2 \leq \lambda_1 < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.2.20}$$

and

$$\mathcal{Y}_{\text{line}}^{\text{Sato}} = \begin{cases} |\lambda_2|\, w(\lambda_1,\lambda_2), & \text{if } \lambda_2 < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.2.21}$$

where the function w is defined as:

$$w(\lambda_1,\lambda_2) = \begin{cases} \left(1 + \frac{\lambda_1}{|\lambda_2|}\right)^{\gamma}, & \text{if } \lambda_2 \leq \lambda_1 \leq 0, \\ \left(1 - \alpha\frac{\lambda_1}{|\lambda_2|}\right)^{\gamma}, & \text{if } \frac{|\lambda_2|}{\alpha} > \lambda_1 > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2.22}$$

We use $\gamma = 0.5$ and $\alpha = 0.25$ as suggested in [189].

- **Li's dot and line filters.** Given eigenvalues that are sorted as $|\lambda_1| \leq |\lambda_2|$, selective filters for the enhancement of dot-like nodules and line-like vessels and suppression of other structures in 2D images are defined as [121]:

$$\mathcal{Y}_{\text{dot}}^{\text{Li}} = \begin{cases} \frac{\lambda_1^2}{|\lambda_2|}, & \text{if } \lambda_1 < 0 \text{ and } \lambda_2 < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.2.23}$$

and

$$\mathcal{Y}_{\text{line}}^{\text{Li}} = \begin{cases} |\lambda_2| - |\lambda_1|, & \text{if } \lambda_2 < 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2.24}$$

- **Frangi's vessel enhancement filter.** The idea behind this vessel enhancement filter is that measures of eigenvalue magnitudes and ratio are incorporated into probability-like estimates. The higher the intensity values of the points in the resulting vesselness image, the higher the probability of the presence of a vessel structure. The eigenvalue magnitudes are measured as the Frobenius norm of the Hessian matrix [63]:

$$\mathcal{X}_{\text{m}} = \sqrt{\lambda_1^2 + \lambda_2^2}, \tag{4.2.25}$$

and their ratio is defined as:

$$\mathcal{X}_{\text{r}} = \frac{\lambda_1}{\lambda_2}, \tag{4.2.26}$$

where $|\lambda_1| \leq |\lambda_2|$.

The vesselness image is computed as [63]:

$$\mathcal{Y}_{\text{line}}^{\text{Frangi}} = \begin{cases} 0, & \text{if } \lambda_2 > 0, \\ \left(1 - \exp\left(-\frac{\mathcal{X}_{\text{m}}^2}{2\gamma^2}\right)\right) \exp\left(-\frac{\mathcal{X}_{\text{r}}^2}{2\beta^2}\right), & \text{otherwise,} \end{cases} \tag{4.2.27}$$

where β and γ are parameters used to normalize $\mathcal{X}_{\text{m}}$ and $\mathcal{X}_{\text{r}}$. To guarantee a strong response only when both the magnitude and ratio criteria are prominent, they are combined by multiplication. In the demonstration of Frangi's method below, we will use $\beta = 0.5$ and $\gamma = 0.5 \max\{\mathcal{X}_{\text{m}}(p, \sigma)\}$, where $\mathcal{X}_{\text{m}}$ is computed for all points of the ROI to which the filtering is applied, as suggested in [63].

Figures 4.15 and 4.16 demonstrate the results of the application of the Lorenz, Sato's, Li's, and Frangi's filters to a patient fluoroscopic image shown in Figure 4.8 (a). Among the presented methods, the best combination of the high filter response to the markers and preservation of their shape is obtained in $\mathcal{Y}_{\text{blob}}^{\text{Sato}}$ and $\mathcal{Y}_{\text{line}}^{\text{Frangi}}$, although many false positives, mostly related to anatomical features, are present. Based on demonstrated results, we expect that designing measures that incorporate the

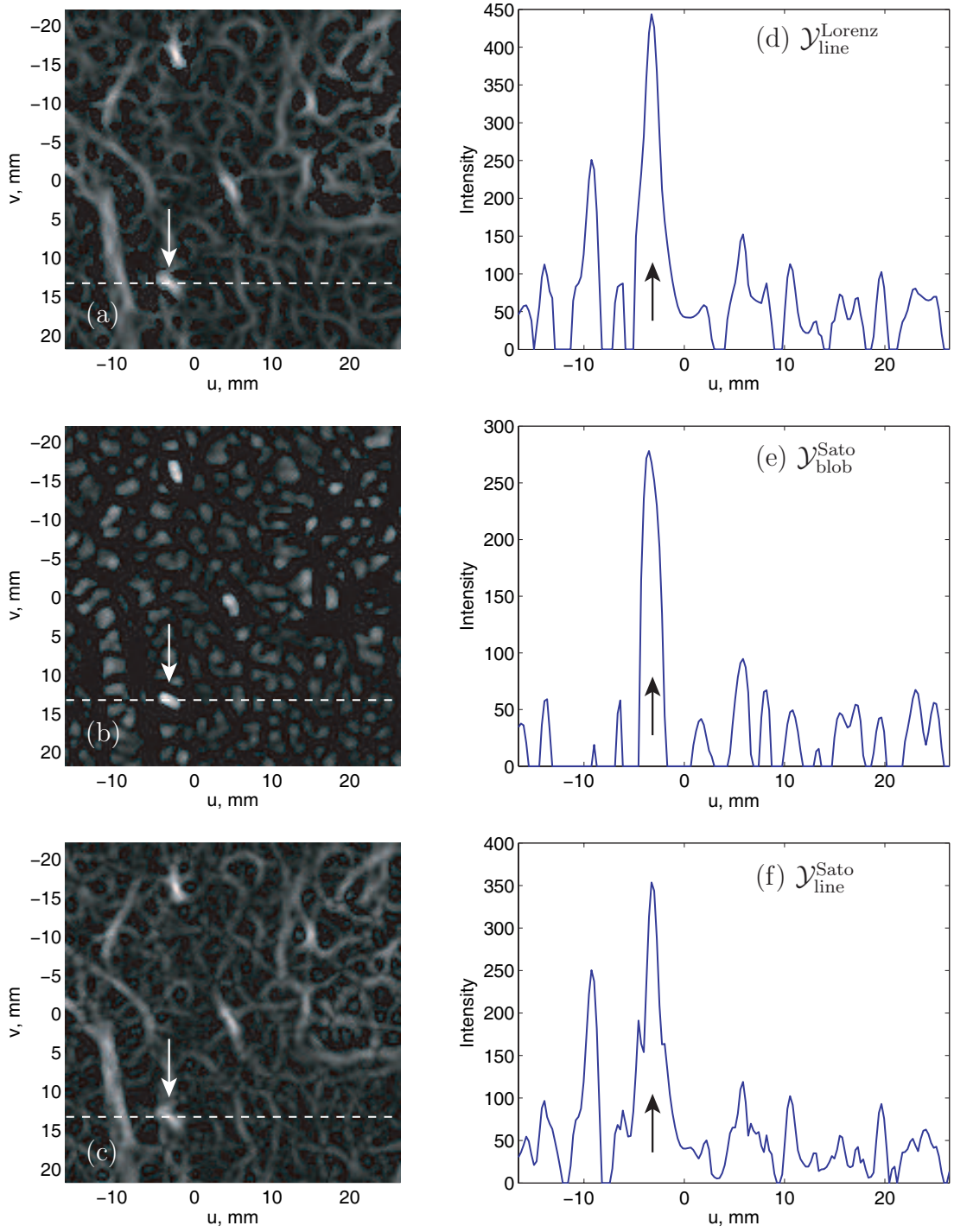


Figure 4.15: The results of the application of Lorenz $\mathcal{Y}^{\text{Lorenz}}_{\text{line}}$ (a), and Sato's $\mathcal{Y}^{\text{Sato}}_{\text{blob}}$ (b) and $\mathcal{Y}^{\text{Sato}}_{\text{line}}$ (c) filters to the image of the prostate patient displayed in Figure 4.8 (a), and their corresponding intensity profiles (d), (e), and (f).

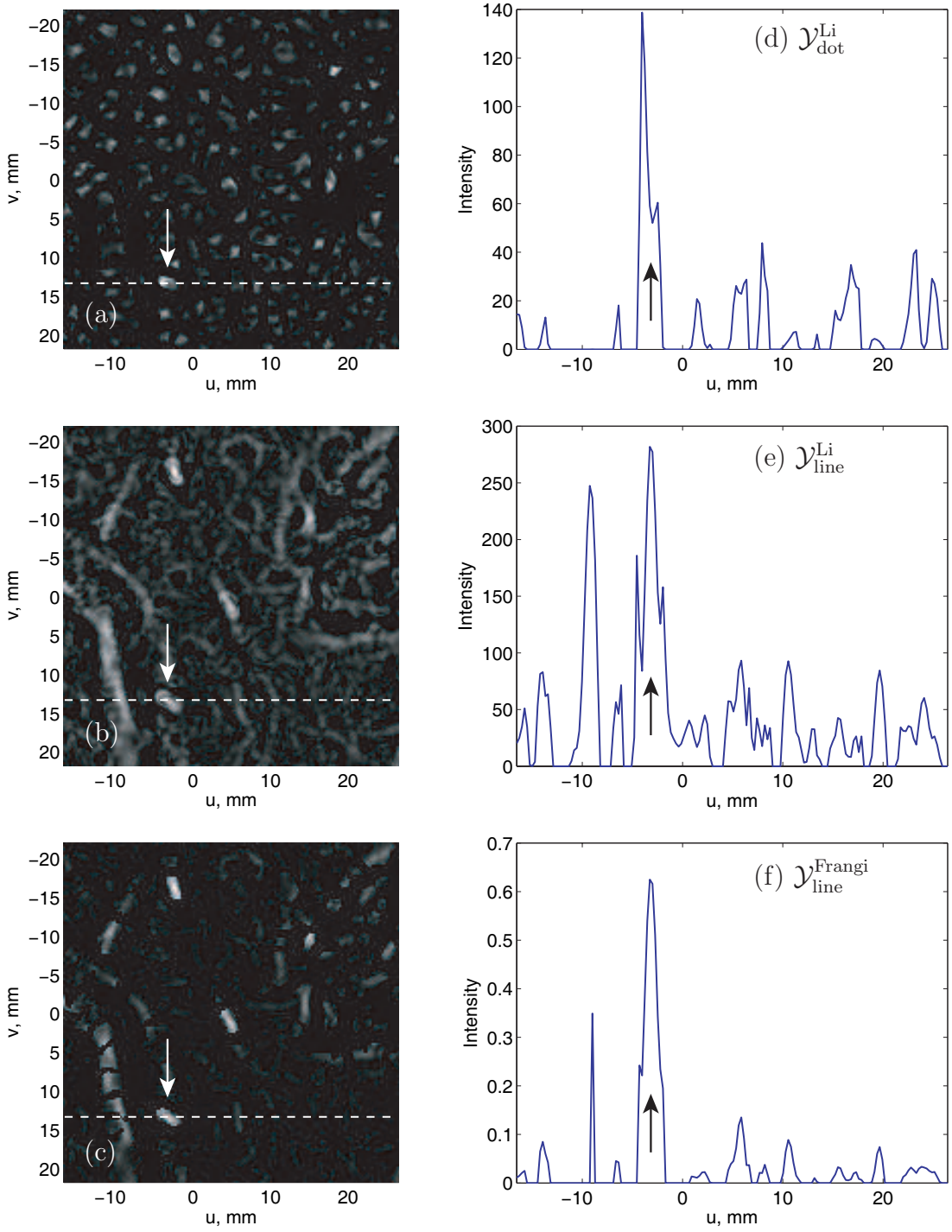


Figure 4.16: The results of the application of Li's $\mathcal{Y}^{\text{Li}}_{\text{dot}}$ (a) and $\mathcal{Y}^{\text{Li}}_{\text{line}}$ (b), and Frangi's $\mathcal{Y}^{\text{Frangi}}_{\text{line}}$ (c) filters to the image of the prostate patient displayed in Figure 4.8 (a), and their corresponding intensity profiles (d), (e), and (f).

probability-like estimates, such as those used in the Frangi's filter, can provide a better suppression of the background and anatomical features in comparison to a direct use of the eigenvalues, such as in the Sato's filter. In Section 4.4, we will examine the application of the principles of the Frangi's approach [63] to our problem, but will design a significantly different filter to that in (4.2.27). The proposed filter is composed of several steps that combine different measures, in order to provide a high response to the markers, preserve their shape, and suppress features in the background. In addition, the proposed marker enhancement filter incorporates a novel contrast enhancement function that acts as a smooth thresholding that further improves the intensity separation between the marker points and the background.

4.3 Computation of Regions of Interest

As will be discussed in Chapters 5 and 6, the computation of the regions of interest (ROIs) in both localization and tracking procedures, $\langle\!\langle \mathbf{L} \rangle\!\rangle$ and $\langle\!\langle \mathbf{T} \rangle\!\rangle$, happens before the call to a filtering procedure $\langle\!\langle \mathbf{F} \rangle\!\rangle$, described in this chapter. While for temporal filtering the results do not depend on the selection of the ROI, and the filtering can be performed either over the ROI or the whole image, the ROI is explicitly used in our marker enhancement filter (MEF). Therefore, we provide the description of this step here, so that the MEF design (Section 4.4) and validation procedures (Section 7.2) are placed in the required context.

Narrowing the focus from the whole fluoroscopic image to a smaller ROI helps to reduce computational burden both for filtering and image registration, and improve reliability: The smaller the image with the sought-for measurements, the lower the chance of failure, which may happen, in particular, in the presence of other features similar in characteristics or causing a similar or higher filter response.

The usage of a smaller ROI within the image proved to be an efficient approach in a number of studies, e.g., [141, 204, 7, 169]. While some authors used manually defined ROIs [7, 169], others designed them as circular or square ROIs of the pre-defined sizes around the expected marker positions computed from the planning CT images [141, 204]. ROIs with a radius or half-side of 2 cm around the projected marker positions were found to be adequate for prostate patients [141, 204], as no displacement larger than 2 cm was observed.

While the application of a fixed ROI size based on the assumptions of the target motion magnitude seems attractive due to its simplicity, we prefer to adaptively compute it, since the filtering procedures in $\langle\!\langle \mathbf{F} \rangle\!\rangle$ prepare the original fluoroscopic image for 2D-3D registration in $\langle\!\langle \mathbf{L} \rangle\!\rangle$ or 2D-2D registration in $\langle\!\langle \mathbf{T} \rangle\!\rangle$. For computational savings and marker enhancement success, the ROI should include the markers but otherwise be as small as possible. However, to be adequate for registration purposes, this ROI should also be large enough to include all possible marker displacements that are allowed by the registration model. The allowable marker displacements are computed based on the geometric transformation used in the registration, 2D or 3D rigid transformation, $\mathcal{T}^{2D}(a, \cdot)$ or $\mathcal{T}^{3D}(a, \cdot)$, respectively, and the bounds on the parameters of these transformations, $a^{\min}$ and $a^{\max}$, where a is a vector of translations and rotations along the coordinate axes. We use symmetric bounds, $a^{\min} = -a^{\max}$, and the size of the bounds for localization and tracking is given in Sections 5.3.1 and 6.1, respectively. The coordinate systems and geometric transformations are discussed in Section 2.4.2 and Appendix B, and the 2D-3D and 2D-2D registration models will be presented in Sections 5.3 and 6.1.

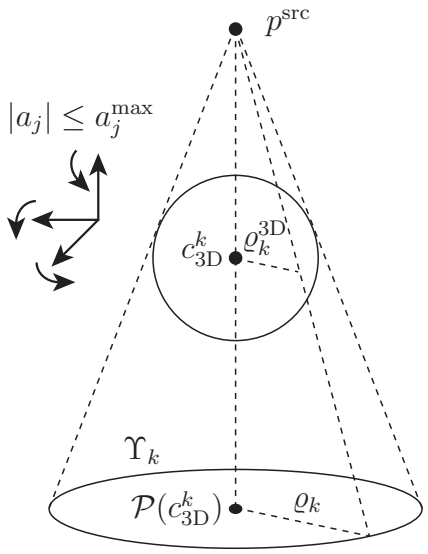


Figure 4.17: Circular marker ROI, Υ_k, computed for localization procedure with $\mathcal{T}^{3\text{D}}$ geometric transformation includes the projection of the 3D marker with centre in $c_{3\text{D}}^k$ and its all possible displacements, maximum of which equals to ϱ_k.

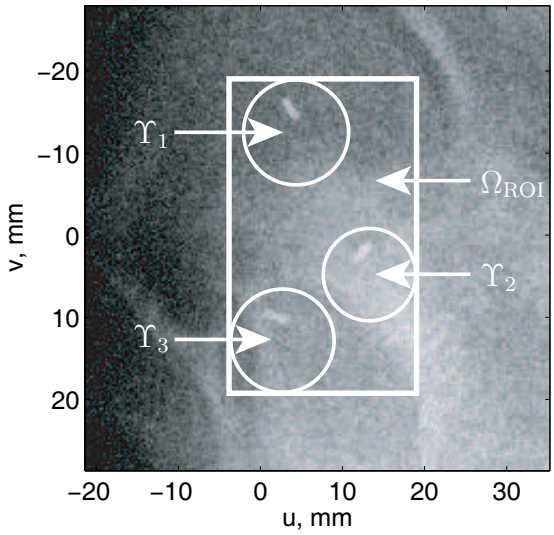


Figure 4.18: Marker ROIs Υ_1, Υ_2, and Υ_3 (*circles*), and registration ROI Ω_{ROI} (*solid rectangle*) overimposed on a fluoroscopic image of a prostate patient.

▲ Υ_k, 2D circular marker ROI

We select a 2D ROI, Υ_k, containing the k^{th} marker and all of its allowable displacements as a circle centered in the predicted marker position with the radius ϱ_k computed as explained below. The predicted position in the localization procedure $\langle\!\langle \mathbf{L} \rangle\!\rangle$ is the point $\hat{c}_{2\text{D}}^k = \mathcal{P}(c_{3\text{D}}^k)$, where $c_{3\text{D}}^k$ is the 3D marker centre found from the CBCT image (Chapter 3), and $\mathcal{P}$ is the 3D to 2D projection operator defined in Equation (B.5.4). For tracking $\langle\!\langle \mathbf{T} \rangle\!\rangle$, the predicted position $\hat{c}_{2\text{D}}^k$ is computed based on the marker positions in the preceding frames, as explained in Sections 6.2 and 6.3. Figure 4.17 demonstrates the principle of computing Υ_k for the localization procedure with $\mathcal{T}^{3\text{D}}$. First, a maximal displacement of the marker in 3D is computed as:

$$\begin{aligned} \varrho_k^{3\text{D}} = \max_a \| c_{3\text{D}}^k - \mathcal{T}^{3\text{D}}(a, c_{3\text{D}}^k) \|, \\ \text{s.t. } |a_j| \leq a_j^{\max}. \end{aligned} \tag{4.3.28}$$

Then, $\varrho_k^{3\text{D}}$ is extended by the half of the marker size, $h/2$, to guarantee the inclusion of the whole marker in the ROI, and scaled to 2D:

$$\varrho_k = \kappa \left(\varrho_k^{3\text{D}} + h/2 \right), \tag{4.3.29}$$

where κ is the 3D to 2D scaling factor for the size of the object that is located in the proximity to the point $c_{3\text{D}}^k \in \mathbb{R}^3$ defined in Equation (B.5.5). For $\mathcal{T}^{2\text{D}}$, in both

localization and tracking, the value for ϱ_k is computed as:

$$\begin{aligned} \varrho_k = \max_a \|\hat{c}^k_{2\mathrm{D}} - \mathcal{T}^{2\mathrm{D}}(a, \hat{c}^k_{2\mathrm{D}})\|, \\ \text{s.t. } |a_j| \leq a_j^{\max}, \end{aligned} \tag{4.3.30}$$

where $\hat{c}^k_{2\mathrm{D}}$ is defined above. Then, the circular marker ROI is computed as:

$$\Upsilon_k = \left\{ q \in \mathbb{R}^2 \mid \|\hat{c}^k_{2\mathrm{D}} - q\| \leq \varrho_k \right\}. \tag{4.3.31}$$

An example of the marker ROIs superimposed on a fluoroscopic image of a prostate patient is shown in Figure 4.18. ■

▲ Ω_{ROI}, **2D rectangular registration ROI**

In addition to Υ_k, we also define a rectangular registration ROI, as in many cases it is more convenient to perform computations over a rectangular ROI that includes all markers, as it translates easily into an array or matrix data structure in the software. We define the registration ROI as:

$$\Omega_{\mathrm{ROI}} = \left\{ q \in \mathbb{R}^2 \mid \min \left\{ \cup_{k=1}^n \Upsilon_k \right\}_{u,v} \leq q_{u,v} \leq \max \left\{ \cup_{k=1}^n \Upsilon_k \right\}_{u,v} \right\}. \tag{4.3.32}$$

In this chapter, we use the notation Ω_{ROI} to refer to both localization and tracking registration ROIs, that are denoted by Ω_{LOC} and Ω_{TR} in Chapters 5 and 6, respectively. Figure 4.18 demonstrates the marker and registration ROIs superimposed on a fluoroscopic image of a prostate patient. The implementation details of the ROI computation are summarized in Algorithm D.5. ■

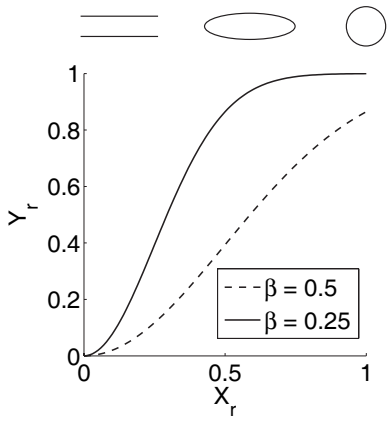


Figure 4.19: The use of $\beta = 0.25$ instead of $\beta = 0.5$ in $\mathcal{Y}_{\mathrm{r}} = 1 - \exp\left(-\mathcal{X}_{\mathrm{r}}^2/(2\beta^2)\right)$ results in high filter response not only in circular blobs, where $\mathcal{X}_{\mathrm{r}} = \lambda_1/\lambda_2 \approx 1$, but also in elliptical blobs, where $\mathcal{X}_{\mathrm{r}} < 1$. The figures above the graph schematically show the shapes that correspond to different values of $\mathcal{X}_{\mathrm{r}}$.

4.4 Novel Marker Enhancement Filter (MEF)

As reported in Section 4.2, the existing methods for feature enhancement that we have evaluated did not provide sufficient marker amplification and background suppression. In this section, we propose the design of a novel marker enhancement filter (MEF) based on the scale-space approach, and inspired by the Frangi's vessel enhancement filter [63].

4.4.1 Principles of MEF Design

Analyzing the design principles of the Frangi's vessel enhancement filter [63] discussed in Section 4.2.2, we make the following two observations:

- Vessels are characterized by a much larger length than width. This means that points belonging to bright vessel measurements on a dark background should possess eigenvalues with the following properties: Having $|\lambda_1| \leq |\lambda_2|$, it should hold that $\lambda_1 \approx 0$ and $\lambda_2 \ll 0$.[9] In contrast, the dimensions of the cylindrical fiducial markers, namely their diameter and length, are not very different. Consequently, if the probability-like measure that includes the eigenvalue ratio, $\mathcal{X}_{\mathrm{r}}$, defined in Equation (4.2.26), is used for marker enhancement, it should attain its maximum values when $\mathcal{X}_{\mathrm{r}}$ is around 1 but not around 0 as required for vessel

[9]The eigenvalue ratio is inversely related to the ratio of semiaxes of a second order shape formed by the image intensity function as $\lambda_1^2/\lambda_2^2 = a_2/a_1$. For the vessels, the semiaxis a_1 that is oriented along the vessel becomes very large or infinite in length, which means that $\lambda_1 \approx 0$.

enhancement. Hence, we define a *magnitude-and-ratio image* as:

$$\mathcal{Y}_{\text{mr}} = \begin{cases} \left(1 - \exp\left(-\frac{\mathcal{X}_{\text{m}}^2}{2\gamma^2}\right)\right)\left(1 - \exp\left(-\frac{\mathcal{X}_{\text{r}}^2}{2\beta^2}\right)\right), & \text{if } \lambda_1 < 0 \text{ and } \lambda_2 < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.4.33}$$

where $\mathcal{X}_{\text{m}}$ is defined in Equation (4.2.25), and $\gamma = 0.5 \max\{\mathcal{X}_{\text{m}}(p, \sigma) \mid p \in \Omega_{\text{ROI}}\}$, as suggested by Frangi *et al.* [63]. Note that in contrast to vessel and other line enhancement techniques, such as those defined in Equations (4.2.19), (4.2.21), (4.2.24), and (4.2.27), we require both λ_1 and λ_2 to be negative. As for β, having a value of 0.5 suggested in [63] will result in the highest $\mathcal{Y}_{\text{mr}}$ values occurring at the points where $\mathcal{X}_{\text{r}}$ is close to 1, i.e., for the circular blobs. We reduce β to 0.25, a value that was selected empirically based on visual observations of the filtered images, in order to produce high filter response for a range of shapes from circular to elliptical blobs (see Figure 4.19). Although the marker dimensions are known, it is not practical to tune the filter to elliptical blobs with the corresponding semiaxes ratio exclusively, as the markers in fluoroscopic images are often corrupted by the noise and overlaying features.

- Observing that the markers may often have reasonably high contrast in fluoroscopic images, we have also considered a *magnitude-only image* that we define as:

$$\mathcal{Y}_{\text{m}} = \begin{cases} 1 - \exp\left(-\frac{\mathcal{X}_{\text{m}}^2}{2\gamma^2}\right), & \text{if } \lambda_1 < 0 \text{ and } \lambda_2 < 0, \\ 0, & \text{otherwise,} \end{cases} \tag{4.4.34}$$

  where γ is the same as in the magnitude-and-ratio image. Note that both $\mathcal{Y}_{\text{mr}}$ and $\mathcal{Y}_{\text{m}}$ take their intensity values within the $[0, 1]$ range.

Analyzing the effect of $\mathcal{Y}_{\text{mr}}$ and $\mathcal{Y}_{\text{m}}$ on patient and phantom images, where sample images are shown in Figure 4.20, we make the following observations that govern the MEF design:

- The enhanced markers in $\mathcal{Y}_{\text{m}}$ retain good preservation of the shapes of the original unfiltered markers. Although they can appear slightly more round[10] in $\mathcal{Y}_{\text{m}}$ than in the fluoroscopic image, their dimensions and orientations are very similar.
- However, due to the absence of control over the feature shapes in $\mathcal{Y}_{\text{m}}$, it is possible that anatomical features similar in scale to the markers are undesirably amplified. For example, note the long vertical feature in Figure 4.20 (a) encompassed by a *rectangle*.

[10] This is due to the use of 0.7 mm value for the Gaussian scale parameter, σ, which is slightly larger than what is required to have the same width of the markers. The decreasing of σ is not desirable as this may produce non-contiguous markers, which is further discussed in Section 4.4.4.

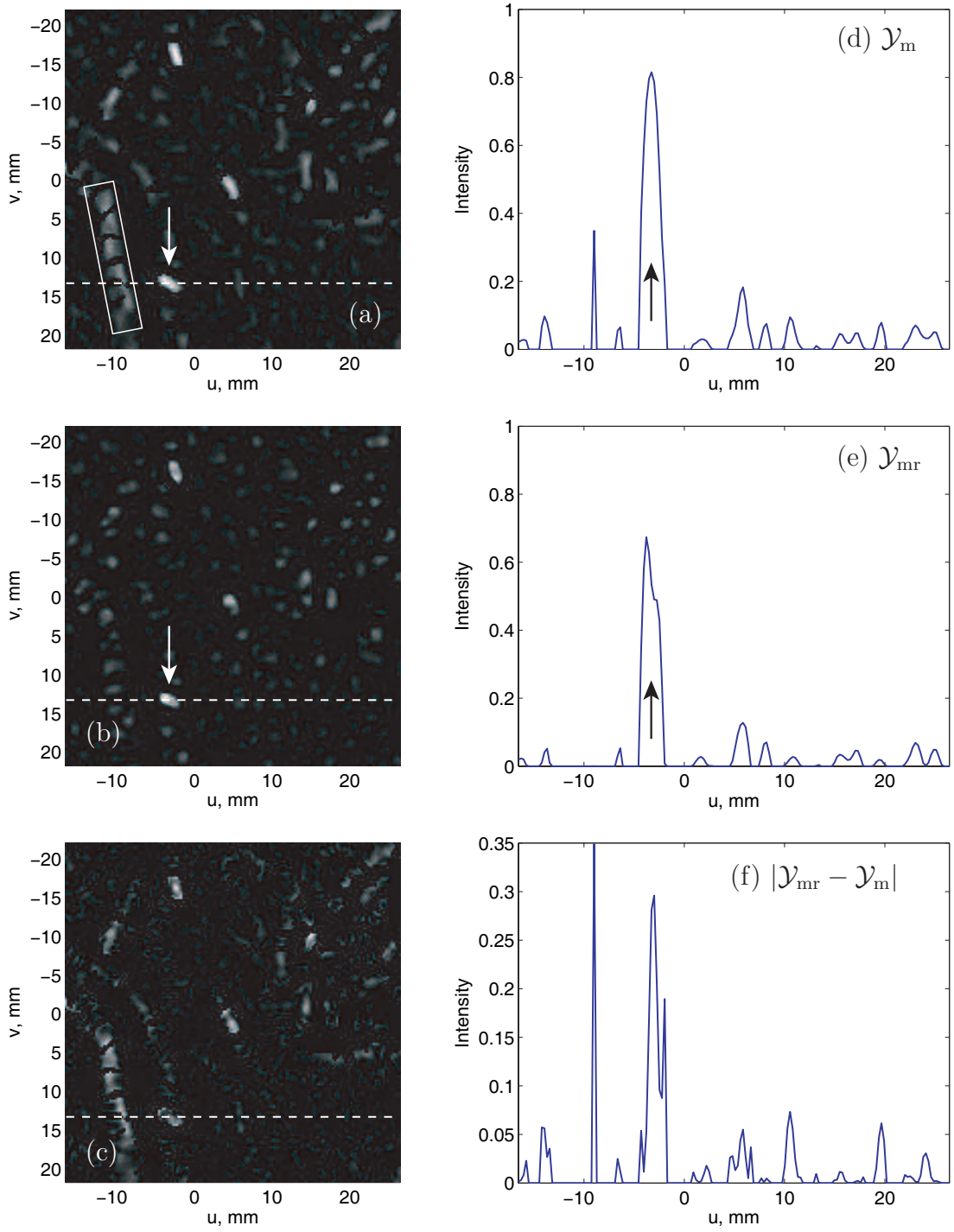


Figure 4.20: The magnitude-only $\mathcal{Y}_{\mathrm{m}}$ (a), magnitude-and-ratio $\mathcal{Y}_{\mathrm{mr}}$ (b) images, their absolute difference (c), and their corresponding intensity profiles (d), (e), and (f). *Rectangle* marks an anatomical feature that was undesirably amplified in $\mathcal{Y}_{\mathrm{m}}$.

- Comparing $\mathcal{Y}_{\text{mr}}$ and $\mathcal{Y}_{\text{m}}$ for different patient and phantom images, we observed that while both of them tend to attain high intensity values in the marker points, many of the high-intensity points belonging to the noise and the anatomical features in one image do not possess high intensities in the other image. For example, Figure 4.20 (c) illustrates the absolute difference between $\mathcal{Y}_{\text{mr}}$ and $\mathcal{Y}_{\text{m}}$. The noticeable difference in the marker points is attributed to the different marker shapes in $\mathcal{Y}_{\text{mr}}$ and $\mathcal{Y}_{\text{m}}$ (almost flat and rectangular in $\mathcal{Y}_{\text{m}}$ versus smaller in length with varying intensity in $\mathcal{Y}_{\text{mr}}$). However, it is easy to see that many of the points belonging to anatomical measurements, such as the long vertical feature to the left of the markers, possess high intensities in $\mathcal{Y}_{\text{m}}$ and low in $\mathcal{Y}_{\text{mr}}$.

Hence, we conclude, that an approach that will provide effective marker enhancement, noise and background elimination, and good marker shape preservation should be based on computing $\mathcal{Y}_{\text{m}}$ values only in the tight regions where the markers can be found, and creating a flat background in all other areas of the image. One of the ways to detect these small regions with high probability of marker presence is to analyze values of $\mathcal{Y}_{\text{mr}}$.

The details of the proposed MEF are presented in Section 4.4.3. To provide a concise discussion in the following sections, the contrast enhancement function used in the MEF to perform smooth separation between the enhanced candidate marker points and the background is defined beforehand in Section 4.4.2.

4.4.2 Contrast Enhancement Function

A common approach for the enhancement of the filtered image is to apply an intensity threshold, which is designed to separate the sought-for markers from the background, e.g., [35, 169, 159]. Let $\mathcal{Y} : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ be a filtered image at a particular scale. Then, a simple *step thresholding* function $\mathcal{C}_{\text{step}} : \mathbb{R} \to \mathbb{R}$ can be defined as:

$$\mathcal{C}_{\text{step}}(\mathcal{Y}(p,\sigma)) = \begin{cases} 0, & \text{if } \mathcal{Y}(p,\sigma) < \vartheta, \\ 1, & \text{otherwise,} \end{cases} \tag{4.4.35}$$

where ϑ is the threshold value. An example of $\mathcal{C}_{\text{step}}$ with $\vartheta = 0.3$ and $0 \leq \mathcal{Y}(p,\sigma) \leq 1$ is shown in Figure 4.21 (*left*). The result of the thresholding operation is a binary image with 1's (white) in the marker points and 0's (black) in the background points. Alternatively, it can be a *grayscale* image where the intensity values above ϑ are kept while the values below ϑ are set to 0:

$$\mathcal{C}_{\text{gray}}(\mathcal{Y}(p,\sigma)) = \begin{cases} 0, & \text{if } \mathcal{Y}(p,\sigma) < \vartheta, \\ \mathcal{Y}(p,\sigma), & \text{otherwise,} \end{cases} \tag{4.4.36}$$

where an example of $\mathcal{C}_{\text{gray}}$ with $\vartheta = 0.3$ and $0 \leq \mathcal{Y}(p,\sigma) \leq 1$ is shown in Figure 4.21 (*centre*). In the MEF, we will also use thresholding. However, we prefer to use a

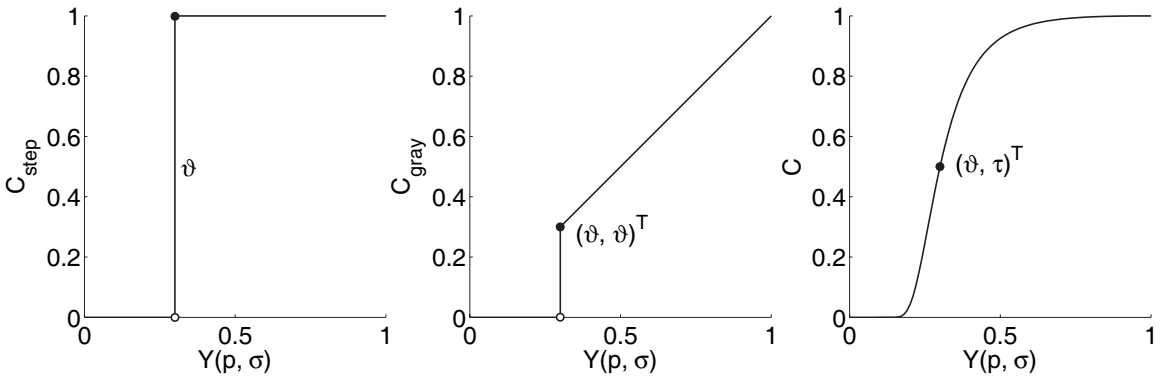


Figure 4.21: Step $\mathcal{C}_{\text{step}}$ (*left*), grayscale $\mathcal{C}_{\text{gray}}$ (*centre*), and smooth $\mathcal{C}$ (*right*) contrast enhancement functions. In this example, $\vartheta = 0.3$. The curve $\mathcal{C}$ (*on the right*) is parameterized by $\tau = 0.5$ and $s = 3$.

smooth threshold instead of $\mathcal{C}_{\text{step}}$ or $\mathcal{C}_{\text{gray}}$, such as shown in Figure 4.21 (*right*). The rationale behind this is that it increases the overall reliability of the method: If some of the markers produce a low filter response, they will be zeroed out by $\mathcal{C}_{\text{step}}$ and $\mathcal{C}_{\text{gray}}$. On the contrary, these measurements will only be partially suppressed, not zeroed, by the smooth threshold function, and can still provide useful information during subsequent processing, such as image registration.

In the proposed threshold function $\mathcal{C}$, separation into the marker and non-marker points is done based on prior knowledge of fiducial marker size, shape and number rather than on the varying intensity characteristics of different images. It is assumed that the majority of the marker points have higher filter response than the non-marker points.

The smooth threshold function should possess the following properties (see Figure 4.21 (*right*) as a guideline):

- $\mathcal{C}(0) = 0$;
- For all $\mathcal{Y}(p, \sigma) \geq \vartheta$, it should hold that $\mathcal{C}(\mathcal{Y}(p, \sigma)) \geq \tau$. Therefore, if we set ϑ to be the lowest intensity found among the candidate marker points (which is determined as described later in this section), then for each candidate marker point $\mathcal{C}(\mathcal{Y}(p, \sigma)$ is not smaller than τ;
- $\mathcal{C}(\vartheta) = \tau$;
- For all $\mathcal{Y}(p, \sigma) < \vartheta$, the result of thresholding should be less than τ;
- $\mathcal{C}(1) = 1$;
- $\mathcal{C}$ should be a non-decreasing function.

One of the functions that satisfies aforementioned properties is a power function

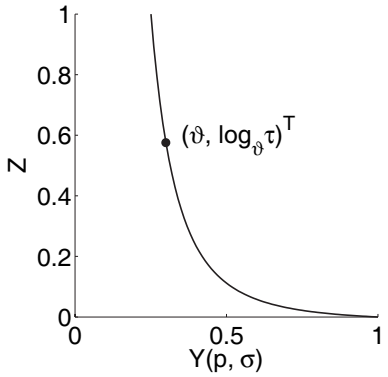


Figure 4.22: Power function $\mathcal{Z}$ used in the smooth contrast enhancement function $\mathcal{C}$. $\mathcal{Z}$ is parameterized by $\vartheta = 0.3$, $\tau = 0.5$, and $s = 3$.

$\mathcal{Z} : \mathbb{R} \to \mathbb{R}$, such that the contrast enhancement function is defined as:

$$\mathcal{C}(\mathcal{Y}(p,\sigma)) = \mathcal{Y}(p,\sigma)^{\mathcal{Z}(\mathcal{Y}(p,\sigma))}, \tag{4.4.37}$$

where, since $0 \leq \mathcal{Y}(p,\sigma) \leq 1$, $\mathcal{Z}$ is a non-increasing function that should satisfy:

$$\begin{cases} \mathcal{Z}(1) = 0, \\ \mathcal{Z}(\vartheta) = \log_\vartheta \tau, \end{cases} \tag{4.4.38}$$

such as shown in Figure 4.22. In particular, $\mathcal{Z}$ can be defined as:

$$\mathcal{Z}(\mathcal{Y}(p,\sigma)) = \frac{a}{\mathcal{Y}(p,\sigma)^s} + b, \tag{4.4.39}$$

where parameter s determines the steepness of $\mathcal{Z}$, and is interpreted as the sharpness of the distinction between the candidate marker points and the background in the final contrast enhancement function $\mathcal{C}$. The function is steeper for larger values of s. Solving (4.4.38) for a and b by substituting $\mathcal{Z}$ defined in (4.4.39) results in the following expression:

$$\mathcal{Z}(\mathcal{Y}(p,\sigma)) = \frac{\vartheta^s \log_\vartheta \tau}{\vartheta^s - 1} \left(1 - \frac{1}{(\mathcal{Y}(p,\sigma))^s}\right). \tag{4.4.40}$$

The remainder of this section explains the choice of the parameters for $\mathcal{Z}$.

▲ τ, **Minimal intensity of the candidate marker points after contrast enhancement**

The value $\tau = 0.5$ is used, so that the intensity values of the marker points occupy the upper half of the image intensity range. ■

▲ s, **Sharpness of distinction**

We assign $s = 3$ as it was found to offer slight improvement for fluoroscopic images compared to $s = 2$. There was no improvement observed for $s > 3$. ■

▲ ϑ, **Marker-based threshold, or the lowest intensity found among the candidate marker points before contrast enhancement**

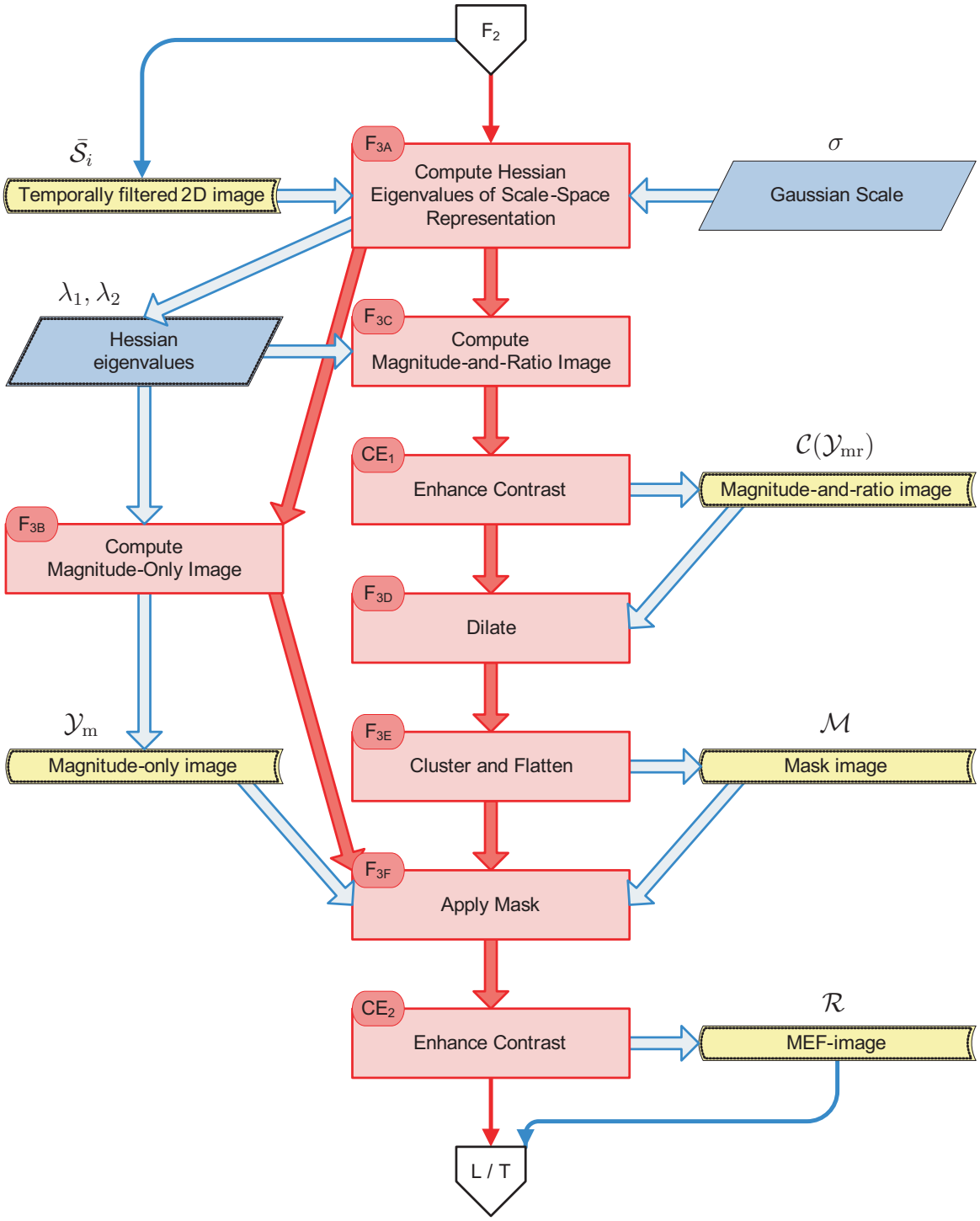
The principle of computing ϑ is similar to that of the model-based segmentation threshold for 3D CBCT images described in detail in Section 3.1.1. For $\mathcal{Z}$ computed on a discrete 2D image, ϑ is the minimal intensity found among m brightest pixels, where m is the estimated number of pixels that constitute the markers. It is found from the number of markers, n, and the number of pixels per marker defined in Equation (4.1.4) in Section 4.1.2. In other potential applications, where the number or the size of the fiducial markers is not known in advance, the MEF can still be used with the intensity-based contrast enhancement, where ϑ is set to some fixed threshold intensity value. For example, $\vartheta = 0.5$ worked reasonably well for our fluoroscopic images. ■

4.4.3 MEF Design Step by Step

The MEF schematic is provided in Flowchart 4.2. This section presents the details of the MEF design, with an example of the MEF step by step application to a patient image in Figure 4.23. Corresponding implementation details are given in Section D.2.

The MEF consists of the following steps:

- $\langle\!\langle \mathbf{F}_{3A} \rangle\!\rangle$: The scale-space representation $\mathcal{L}$ is computed for a temporally filtered fluoroscopic image $\bar{\mathcal{S}}_i$. The scale $\sigma = 0.7$ mm is selected in relation to the marker size (more discussion is provided in Section 4.4.4). Next, the Hessian eigenvalues $\lambda_1(p, \sigma)$ and $\lambda_2(p, \sigma)$ are computed at all points $p \in \Omega_{\text{ROI}}$ of the image $\mathcal{L}(p, \sigma)$. Algorithm D.7 describes implementation details of this step.
- $\langle\!\langle \mathbf{F}_{3B} \rangle\!\rangle$, $\langle\!\langle \mathbf{F}_{3C} \rangle\!\rangle$, and $\langle\!\langle \mathbf{CE}_1 \rangle\!\rangle$: The images $\mathcal{Y}_{\text{m}}$ and $\mathcal{Y}_{\text{mr}}$ are computed as defined in Equations (4.4.34) and (4.4.33), respectively. Then, the image $\mathcal{Y}_{\text{mr}}$ undergoes contrast enhancement with the function $\mathcal{C}$ defined in Equation (4.4.37) and detailed in Algorithm D.8. It is performed to suppress high intensities of some of the non-marker points to improve filtering outcome and reduce computational cost of the next steps, in particular, flattening.
- $\langle\!\langle \mathbf{F}_{3D} \rangle\!\rangle$ and $\langle\!\langle \mathbf{F}_{3E} \rangle\!\rangle$: Computation of the mask $\mathcal{M}$. The markers in $\mathcal{C}(\mathcal{Y}_{\text{mr}})$ often appear smaller than in $\mathcal{Y}_{\text{m}}$ and $\bar{\mathcal{S}}_i$ (see Figure 4.23). To create an effective mask, the high-intensity regions in $\mathcal{C}(\mathcal{Y}_{\text{mr}})$ are enlarged by *morphological dilation* [73], described in Section 4.2.1. As a structuring element, we used a circle of a radius $h/2$, where h is the marker size.



Flowchart 4.2: Block $\langle\langle \mathbf{F}_2 \rangle\rangle$: The schematic of the marker enhancement filter.

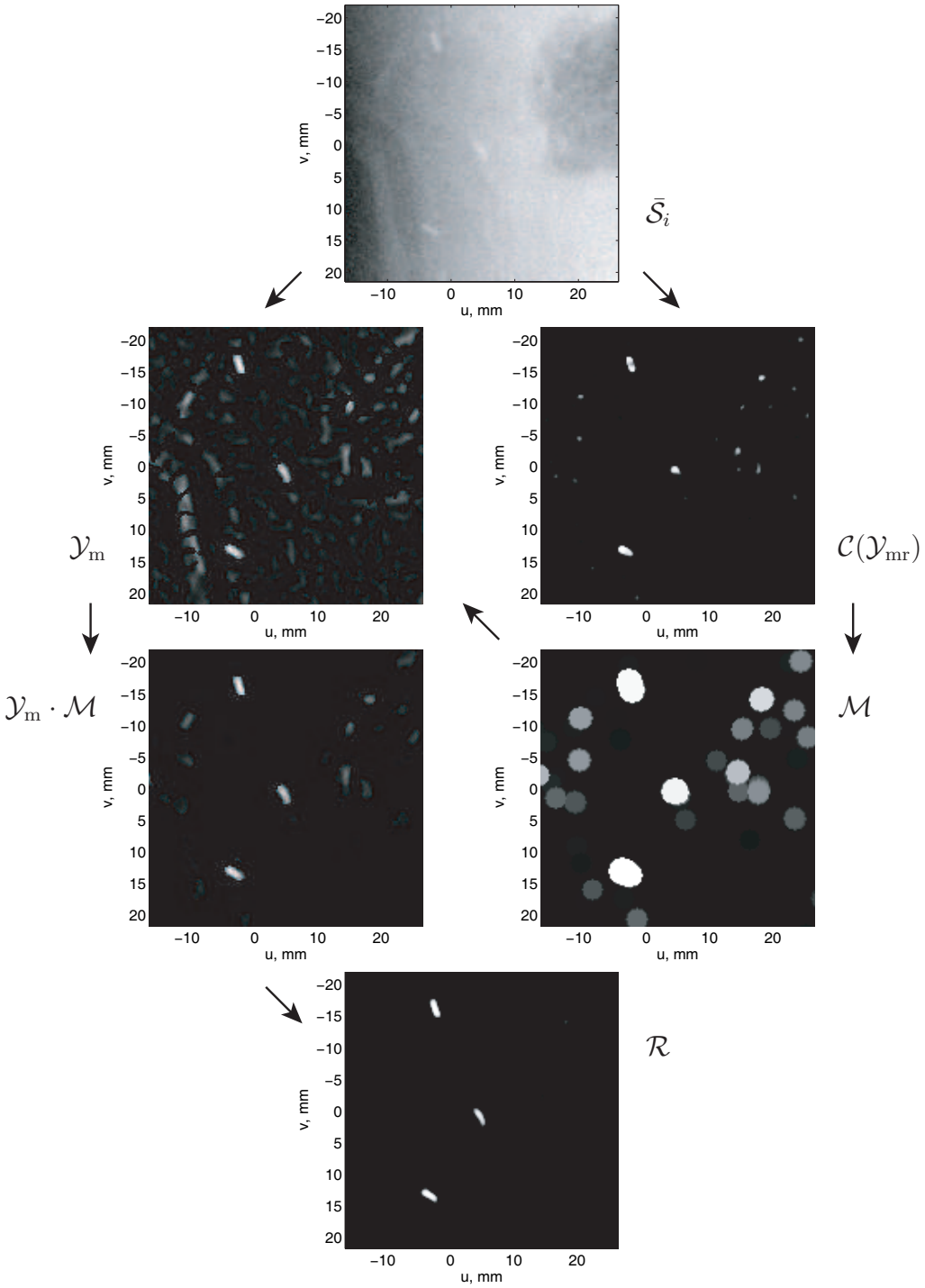


Figure 4.23: Step by step application of the marker enhancement filter (MEF) to a patient image. The magnitude-only $\mathcal{Y}_\mathrm{m}$ and the contrast-enhanced magnitude-and-ratio image $\mathcal{C}(\mathcal{Y}_\mathrm{mr})$ are created from the temporally filtered fluoroscopic image $\bar{\mathcal{S}}_i$. Then, the high-intensity regions in $\mathcal{C}(\mathcal{Y}_\mathrm{mr})$ are enlarged and flattened to create a mask $\mathcal{M}$, which is applied to $\mathcal{Y}_\mathrm{m}$. In the end, the MEF-image $\mathcal{R}$ is produced from $\mathcal{Y}_\mathrm{m} \cdot \mathcal{M}$ by the final application of the contrast enhancement function $\mathcal{C}$.

Next, the mask $\mathcal{M}$ is created by *flattening* the contiguous regions in the dilated image. This is done by assigning the maximal intensity found within each region to all of its points, where the intensity is related to the probability that the region encloses a marker. The flattening is performed in order to exclude any distortion of $\mathcal{Y}_{\mathrm{m}}$ intensities during the subsequent mask application, except for the scaling. To group the points, the *hierarchical clustering* technique is used, which was described in Section 3.1.2. According to this method, a point is added to a cluster if the shortest distance between it and the cluster is less than $1.5\,\delta_{\mathrm{px}}$, where δ_{px} is a pixel size. Hence, points of the same cluster are either connected by their sides (δ_{px} distance in between) or by their corners ($\delta_{\mathrm{px}}\sqrt{2}$ distance).

Clustering and flattening is only performed for the points whose intensity is greater or equal to the intensity threshold ϑ, defined in Section 4.4.2, i.e., the points that most probably belong to the regions enclosing the markers, both to reduce computational cost and to prevent amplification of non-marker features. The implementation details of the computation of $\mathcal{M}$ are provided in Algorithm D.9.

- $\langle\!\langle \mathbf{F}_{3\mathrm{F}} \rangle\!\rangle$ and $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$: The mask $\mathcal{M}$ is applied to $\mathcal{Y}_{\mathrm{m}}$, and the final MEF-image is produced after another contrast enhancement step with the function $\mathcal{C}$:

$$\mathcal{R} = \mathcal{C}(\mathcal{Y}_{\mathrm{m}} \cdot \mathcal{M}). \tag{4.4.41}$$

In order to further improve the outcome of the marker enhancement, all intensities in $\mathcal{R}$ outside of the circular marker ROIs Υ_k, defined in Section 4.3, are zeroed out. The final contrast enhancement step $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$, however, is optional, and depends on the images at hand and the purposes of the project. More details are provided in Section 4.5.

4.4.4 MEF Parameters

While the selection of β (relates to the shape of the features we wish to amplify with the MEF) and σ (the Gaussian scale) has been briefly covered earlier (Sections 4.4.1 and 4.2.2, respectively), this section offers final notes on parameter selection.

Although the use of $\beta = 0.25$ extends the range of the shapes that are amplified from circular to elongated blobs in comparison to $\beta = 0.5$ proposed by Frangi *et al.* [63], typical marker shapes correspond to even smaller values of β. On the other hand, using a smaller β, such as 0.1 or below, results in a very small contribution from the ratio measure, $\mathcal{X}_{\mathrm{r}}$, to the resulting image $\mathcal{Y}_{\mathrm{mr}}$, in which case $\mathcal{Y}_{\mathrm{mr}}$ becomes similar to $\mathcal{Y}_{\mathrm{m}}$, which has high intensity values in some of the anatomical features. Hence, in the proposed MEF design, the combination of $\sigma = 0.7$ mm and $\beta = 0.25$ produces high

$\mathcal{Y}_{\text{mr}}$ values at one or both ends of the markers, as can be seen in Figures 4.20 and 4.23. Then, these small patches of high-intensity points are extended by dilation into the regions that are large enough to include the whole marker.

The Gaussian scale $\sigma = 0.7$ mm produces the best results for the fiducial markers 3×0.8 mm in size. The use of smaller values, such as $\sigma = 0.4$ mm (Figure 4.24, *left column*), does not sufficiently suppress noise and tends to create chopped and non-contiguous markers. Larger values, such as $\sigma = 1$ mm (Figure 4.24, *middle column*), generally work well but distort marker shape by generating marker measurements that are smoothed out too much (rounded). Finally, values such as $\sigma = 1.5$ mm (Figure 4.24, *right column*), tend to amplify anatomical features. It can be said that σ is adjusted to the marker diameter rather than the marker length because it helps to preserve the shape, and also helps to efficiently suppress the background. The noise component of the background is typically on smaller scales while many of the anatomical features are on larger scales.

4.4.5 Magnitude-Only Filter (MOF)

As was mentioned in Section 4.4.1, we believe that effective marker enhancement, noise and background suppression, and marker shape preservation can be achieved by computing the intensity values of the magnitude-only image, $\mathcal{Y}_{\text{m}}$, in the regions where the markers are likely to be found, and enforcing a flat background outside of these regions. In the MEF, we found these regions with the help of the magnitude-and-ratio image, $\mathcal{Y}_{\text{mr}}$. In this form, the MEF was designed to be used for 2D marker localization, where the displacement between the markers in the 3D CBCT and 2D fluoroscopic images can be significant. During tracking, on the other hand, the marker displacement between the frames is expected to be significantly smaller. Therefore, the regions can be defined around the predicted marker positions, which can be computed by using information from preceding images of the sequence as will be explained in Section 6.2. Hence, for tracking we define a variation on the MEF, a magnitude-only filter (MOF). It consists of steps $\langle\!\langle \mathbf{F}_{3\text{A}} \rangle\!\rangle$, $\langle\!\langle \mathbf{F}_{3\text{B}} \rangle\!\rangle$, and $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$ described in Section 4.4.3. Although the full MEF can be used in tracking as well, this simplification allows the reduction in computational cost, mainly due to the removal of the clustering and flattening step $\langle\!\langle \mathbf{F}_{3\text{E}} \rangle\!\rangle$.

4.5 Discussion

By analyzing the results of the MEF application to several testing data sets reported in Section 7.2.3 and comparing its performance to a number of other filters in Section 7.2.4, we concluded that the MEF is an effective choice for the marker enhance-

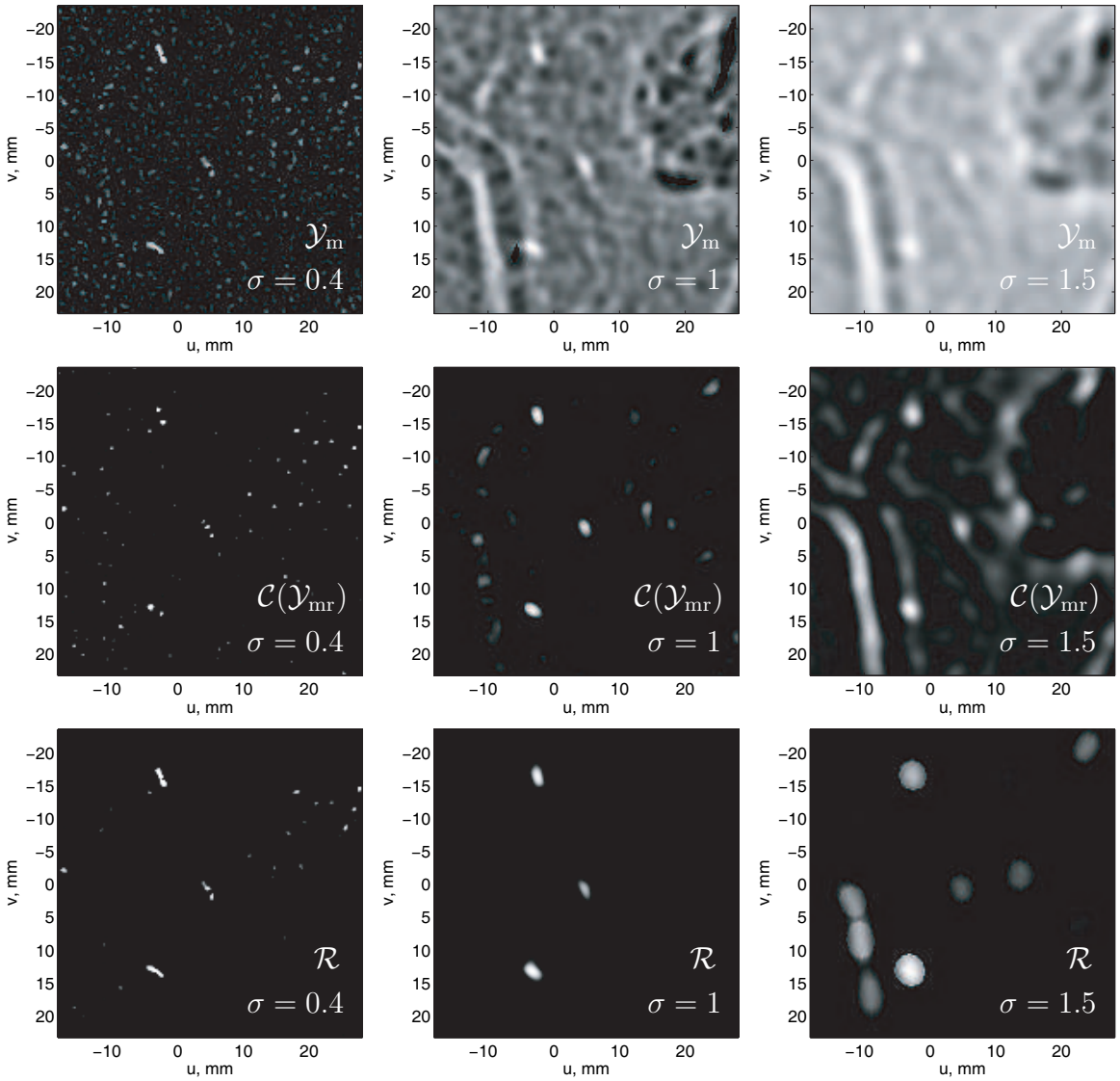


Figure 4.24: Images $\mathcal{Y}_{\mathrm{m}}$, $\mathcal{C}(\mathcal{Y}_{\mathrm{mr}})$ and $\mathcal{R}$ produced with the Gaussian scales $\sigma = 0.4$ mm (*left column*), 1 mm (*middle column*), and 1.5 mm (*right column*).

ment task. In fact, to the best of our knowledge, it is the best among the available filters.

Although the MEF was designed with a particular application in mind, it can be used in a variety of marker localization and tracking applications. The main advantages of the proposed filter over other filters are its marker enhancement and shape preservation properties, and its successful background suppression. No preliminary information about the marker orientation is required. The quantitative results presented in Sections 7.2.3 and 7.2.4 suggest that the use of temporal filtering, in particular, the Dolph-Chebyshev filter, prior to the MEF application improves the filtering outcome. From our experience, temporal filtering also helps to reduce the MEF processing time, as the overhead associated with opening multiple image files for temporal filtering is significantly smaller than the time savings resulting from the ability to process a smoother image. The savings come from the decrease in the number of high-intensity pixels that should participate in the grouping and flattening procedures at step $\langle\!\langle \mathbf{F}_{3E} \rangle\!\rangle$. In addition to a visibly better preservation of the marker shapes and suppression of the background, the MEF also performs similarly or better than other filters in terms of quantitative assessment of marker enhancement. Section 7.2.4 presents comparative results between the MEF and other filters reported in the literature, such as the MEK [160], LoG^2 [127], Sato's blob filter [189], and Frangi's line filter [63] reviewed in Section 4.2.

The main limitation of the MEF comes from the assumption that among all the features observed at the selected scale σ, the markers should possess the highest contrast. While a couple of features similar to the markers in size and contrast typically do not jeopardize successful marker enhancement and subsequent 2D marker localization or tracking, especially considering that in the proposed application they are performed with the use of image registration, the presence of a higher-contrast feature with one of the dimensions corresponding to σ and the other much larger is not desirable.[11]

While the use of the contrast enhancement as the final step $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$ of the MEF or MOF generally helps to amplify the markers, and is especially useful if the image contains features causing a higher filter response than the markers (such as an example with a portal image provided in Section 7.2.4), in some cases the use of the CEF is not necessary (though it is still used at the step $\langle\!\langle \mathbf{CE}_1 \rangle\!\rangle$). From our experience with the lateral images of the prostate patients, one or sometimes several markers can be overlayed by anatomical features. In this case, these markers possess

[11] Ideally, the magnitude of the eigenvalue that corresponds to the feature's lengthwise direction should be small, which makes both $\mathcal{X}_m$ and $\mathcal{X}_r$ small, and as such, the points of the feature should not be amplified by the MEF. However, it is common that the intensity along the feature varies, which makes the filter perceive it as a sequence of blob-like structures.

very low contrast, and can further be suppressed by the final application of the CEF. We observed that the performance of the localization and tracking procedures in application to our patient testing data set was better when the step $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$ was excluded even though the images appeared to be of a lower quality (higher noise and more anatomical features present). In the end, the decision to use the CEF at the step $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$ should be based on knowledge of data and purposes of the processing.

The design and validation of the MEF were reported in [173]. Appendix D.2 provides implementation details for the filtering block $\langle\!\langle \mathbf{F} \rangle\!\rangle$. Temporal filtering procedure $\langle\!\langle \mathbf{F}_1 \rangle\!\rangle$ is summarized in Algorithm D.4, computation of the ROI $\langle\!\langle \mathbf{F}_2 \rangle\!\rangle$ in Algorithm D.5, and the details of the MEF implementation are given in Algorithms D.6–D.9.

Chapter 5

Marker Localization in 2D Space

This chapter presents methods for marker localization in 2D space, corresponding to block $\langle\!\langle \mathbf{L} \rangle\!\rangle$ of the scheme in Flowchart 2.1. Flowchart 5.1 provides details of this block. The localization is performed by the means of *2D-3D image registration* between the MEF-image $\mathcal{R}$ produced in block $\langle\!\langle \mathbf{F} \rangle\!\rangle$ (Chapter 4) and a 2D template image $\mathcal{W}$ formed from the 3D marker sets M_k or models $\mathrm{M}_k^{\mathrm{mod}}$ computed in block $\langle\!\langle \mathbf{S} \rangle\!\rangle$ (Chapter 3).

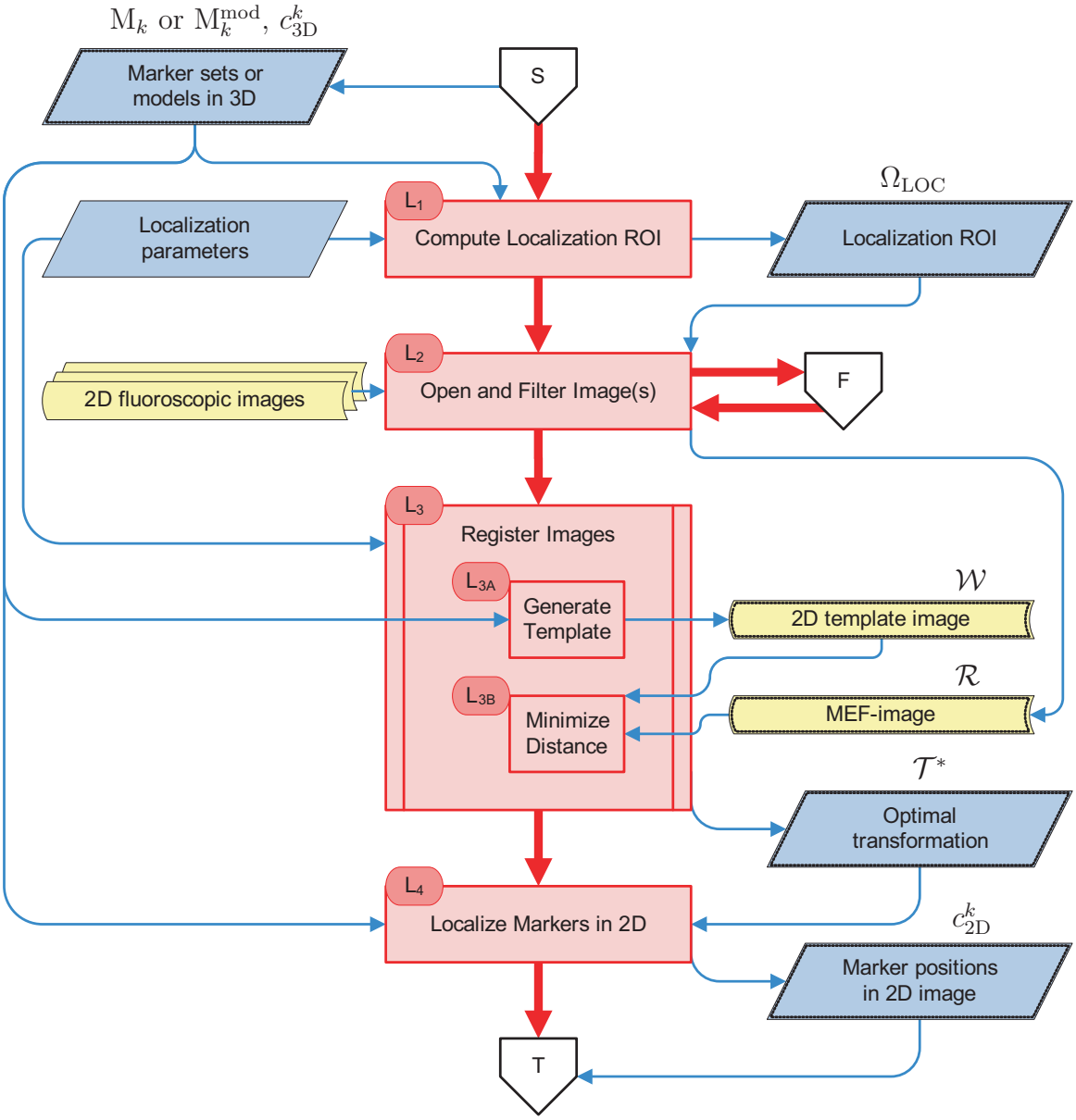
Medical image registration techniques and applications are discussed in a number of books [149, 150, 19, 57, 75] and surveys [143, 69, 83, 138, 81, 119, 145, 34]. At a high level, the image registration problem can be stated as follows: Given two images,[1] $\mathcal{A} : \mathbb{R}^{d_\mathcal{A}} \to \mathbb{R}$ (called the *template*, or *moving image*) and $\mathcal{B} : \mathbb{R}^d \to \mathbb{R}$ (called the *reference*, or *fixed image*) of dimensions $d_\mathcal{A} \in \mathbb{N}$ and $d \in \mathbb{N}$, respectively, we aim to find an *optimal geometric transformation* $\mathcal{T}$ that alters $\mathcal{A}$ to become *similar* to $\mathcal{B}$. The dissimilarity between the *transformed template* $\mathcal{A}_\mathcal{T} : \mathbb{R}^d \to \mathbb{R}$ and the reference $\mathcal{B}$ is quantified by some *distance measure* function $\mathcal{D}$ that is to be *minimized*:

$$\mathcal{T}^* = \arg\min_{\mathcal{T}} \mathcal{D}\left(\mathcal{A}_\mathcal{T}, \mathcal{B}\right). \tag{5.0.1}$$

We assume that function $\mathcal{D}$ returns a real value, and that the minimizer exists. The transformation $\mathcal{T}$ includes the dimension transformation from $d_\mathcal{A}$ to d if $d_\mathcal{A} \neq d$, such as in the case of 2D-3D image registration. The optimization problem can include constraints, bounds on parameters, and regularization on $\mathcal{T}$. The computation of $\mathcal{D}$ is often restricted to some region of interest $\Omega_{\mathrm{LOC}} \subset \mathbb{R}^d$ within images. We will discuss our particular registration problem in more detail in Section 5.3.

Flowchart 5.1 provides a scheme of our proposed 2D marker localization procedure. The 3D positions of the marker sets or models, $c_{3\mathrm{D}}^k$, and the assumptions about the maximal displacement between the markers in the CBCT and fluoroscopic

[1] Although more than two images may be subject to simultaneous registration, in this thesis we only consider two images at a time, as it suffices our purposes.



Flowchart 5.1: Block ⟪**L**⟫: Marker localization in 2D by 2D-3D image registration between a 2D template image formed from the 3D marker sets or models computed in block ⟪**S**⟫ and the MEF-image produced in block ⟪**F**⟫.

images are used to compute a localization ROI, Ω_{LOC}, at step $\langle\!\langle \mathbf{L}_1 \rangle\!\rangle$. At step $\langle\!\langle \mathbf{L}_2 \rangle\!\rangle$, the fluoroscopic images are loaded, and the control is passed to block $\langle\!\langle \mathbf{F} \rangle\!\rangle$, where temporal and marker enhancement filtering takes place, returning a MEF-image $\mathcal{R}$, that is defined on Ω_{LOC}. Step $\langle\!\langle \mathbf{L}_3 \rangle\!\rangle$ corresponds to 2D-3D image registration. In the process of registration, we seek an optimal transformation $\mathcal{T}^*$ that solves a problem in the form of (5.0.1). The optimal transformation aligns the markers in the 2D template image $\mathcal{W}$, produced from the 3D marker sets or models, M_k or M_k^{mod}, respectively, found in the 3D CBCT in block $\langle\!\langle \mathbf{S} \rangle\!\rangle$, to the markers in $\mathcal{R}$. Possible constraints, bounds on variables, optimization stopping criteria, and other settings are collectively referred to as *localization parameters* in Flowchart 5.1. The template image generation methods and the distance measures used at step $\langle\!\langle \mathbf{L}_3 \rangle\!\rangle$ can be treated as self-contained topics. Therefore, for the clarity of narration, we discuss them prior to defining and discussing the solution to the optimization problem, and denote them as steps $\langle\!\langle \mathbf{L}_{3\text{A}} \rangle\!\rangle$ and $\langle\!\langle \mathbf{L}_{3\text{B}} \rangle\!\rangle$ in Flowchart 5.1. Once $\mathcal{T}^*$ is found, it is applied to the 3D marker positions c_{3D}^k at step $\langle\!\langle \mathbf{L}_4 \rangle\!\rangle$ in order to obtain the 2D marker positions c_{2D}^k.

Although many references are available on the image registration framework and its components [149, 150, 19, 57, 75, 143, 69, 83, 138, 81, 119, 145, 34], the specifics of each application may require reasoning and experiments to select the most appropriate combination of component techniques. In our case, we did not want to use any ready-made configuration offered in the literature as we perform registration between unique types of images: Although the initial imaging modalities are common, we use novel processing techniques, such as the MEF described in Section 4.4, and template image generation approaches that will be discussed in Section 5.1. Therefore, this chapter contains reviews of common techniques used in image registration, our justifications based on knowledge about marker localization problem, and experiments on our data sets that help to make choices for particular components of the registration problem.

After Section 5.1, in which we propose three methods for template image generation that are designed specifically for the markers, we proceed by selecting an appropriate distance measure among the commonly used functions based on experiment results on our data sets in Section 5.2. Further, Section 5.3 provides details on our image registration problem formulation. Based on the literature and experiments on our data sets, we select a geometric transformation function in Sections 5.3.1 and 5.3.2. To perform the minimization, we use one of the standard techniques. However, as our minimization problem is generally non-convex, we perform a thorough evaluation of techniques and their combinations to select an appropriate starting point in Section 5.3.3. The overall performance of the 2D marker localization procedure on our testing data sets is used to select the kernel width for the Dolph-Chebyshev

temporal filter in Section 5.4. The chapter is concluded by a brief review of failure detection and prevention techniques in Section 5.5, and a discussion of the results in Section 5.6.

5.1 Template Image Generation

This section describes our novel marker specific template image generation methods. It is referred to as step $\langle\langle \mathbf{L}_{3A} \rangle\rangle$ in Flowchart 5.1.

As was discussed in Section 1.2, a conventional way to perform 2D-3D image registration between a 3D CBCT and a 2D fluoroscopic or x-ray image is to register the 2D image to a template image synthesized from the 3D CBCT, a *digitally reconstructed radiograph* (DRR). In general, creating a DRR is a computationally expensive procedure, especially if it has to be computed multiple times as a part of an iterative optimization scheme, which is often the case in image registration, and is indeed the case in our 2D marker localization procedure (Section 5.3.3).

Unfortunately, high computational cost is not the only disadvantage of the conventional DRR method if used in our marker localization procedure. As the DRR simulates an x-ray projection image, it is naturally prone to the same issues that arise in projection. In particular, in addition to the markers, the DRR will also include the "projections" of multiple anatomical features present in the CBCT image. This will worsen the CNR of the markers and make registration between the markers in the MEF-image and in the generated template extremely challenging. As the fiducial markers are the indicators of the target position, we are interested to align the markers in the two images rather than any other structures.

In the fluoroscopic image, we have overcome these problems by applying temporal and MEF filtering (Chapter 4). To ease the computational burden, and to perform registration based mainly on the markers, we propose to use the techniques described in this chapter to create a *selective DRR* image that is formed from the 3D marker sets or models computed from the 3D CBCT image as was explained in Chapter 3. This approach allows us to create two images, the MEF-image and the 2D template, that have very similar intensity characteristics; both display bright markers on a dark background (given that the segmentation and filtering procedures described in Chapters 3 and 4 were successful).

In the remainder of this introduction, we provide a brief overview of the methods for DRR generation reported in the literature, discuss our novel marker specific selective DRR approaches, and how they are related to the existing techniques. The previously developed methods include raytracing [201, 120, 90], shear-warp factorization [115, 42], attenuation fields [185, 181], splat rendering [28, 232], and point based

methods [7, 197].

The first proposed method is the *marker model DRR*, which is based on the principle of raytracing [201], that is, the intensities in the template image are computed by accumulating the intensities in the 3D imaging data encountered as the ray is traced from the x-ray source to the detector. Our approach differs from the standard method in that we use the 3D marker models instead of the raw data. As the marker models are continuous mathematical representations, this allows for an elegant analytical solution as compared to the process of counting voxels and their intensities used in the original method.

The second method, a *voxel splat DRR*, operates on the 3D marker sets. It is an extension of the original splat rendering technique [232], which was proposed to generate 2D images from 3D data for orthographic views. The method is based on the idea that each voxel's intensity is spread among multiple pixels that lie within the extent of the projected volume of the voxel (footprint). However, in the case of perspective projection, such as used in our problem, the computation may become prohibitively expensive, and some modifications have been proposed to obtain a reasonable image quality in reasonable time [28]. Due to the fact that for the DRR generation we only use voxels that belong to the 3D marker sets, our modification of the original method allows to extend it to perspective projection views while maintaining a reasonable computational cost.

The results presented in Section 7.1.7 suggest that it may not be possible to precisely recover orientations of the fiducial markers from the CBCT images with the voxel sizes of above 1 mm, making it unreasonable to use the marker model DRR. Also, large voxels will produce unreasonably large splats, making it impractical to use the voxel splat DRR. Hence, we propose a third approach, a *marker splat DRR*, inspired by the voxel splat approach. It produces a single splat for each 3D marker in the template image. The size of the splat, which has a circular symmetry, is related to the marker dimensions. Sections 5.1.1, 5.1.2, and 5.1.3 describe the details of the marker model, voxel splat, and marker splat DRR methods, respectively.

5.1.1 Marker Model DRR

In this section, we propose a DRR generation method that uses the 3D marker models, $\mathrm{M}_k^{\mathrm{mod}}$, $k = 1, \ldots, n$, computed in the segmentation block $\langle\!\langle \mathbf{S} \rangle\!\rangle$ (Section 3.4). It is referred to as the marker model DRR (MM DRR). The proposed method is similar to the *raytracing* technique that was designed to produce 2D images from a 3D volume, and, among other applications, was also used for DRR generation [26, 241, 201, 120, 90]. The raytracing method numerically approximates radiographic image formation process, according to which, under a number of simplifications [94, 26], the non-diffracting

homogeneous x-rays attenuate as they pass through the matter. This results in different measurements on the detector elements depending on the properties of the material and the length of the path in the material that the x-rays went through on their way. Each ray contributes to the intensity of some pixel by a value proportional to $\int_{\text{ray}} \mu(p)dp$, where $\mu(p)$ is an attenuation coefficient of the matter at the point $p \in \text{ray} \in \mathbb{R}^3$. The attenuation coefficients are related to the voxel intensities in the CT and CBCT images [94, 26]. For discrete images, the raytracing involves the detection of the voxels $\check{p}^1, \ldots, \check{p}^m$ through which a particular ray passes. Often, to simplify computation, the 2D pixel intensity is computed as a value proportional to $\sum_{i=1}^{m} \mu[\check{p}^i]$, i.e., the path lengths of the ray inside the voxels are neglected, e.g., [26].

Instead of tracing through the discrete data, i.e., through the voxels of the 3D image, we propose to compute the rays through the 3D marker models M_k^{mod}, which are continuous entities. With this approach, high quality projections of the fiducial markers are created that are limited only by the pixel size of the 2D image and do not suffer from the discrete nature of the 3D image, as is often the case with the DRRs created by the conventional raytracing technique.

The following discussion assumes that the imaging system was rotated to its initial position, such as displayed in Figure 2.5, i.e., the x-ray source p^{src} is located on the positive part of the X axis, and the isocentric plane H coincides with the YZ plane. The points are in homogeneous coordinates (Appendix B). The cylindrical marker model can be mathematically described as the intersection of an infinite cylinder of a radius r and two parallel planes that are perpendicular to the cylinder's axis. Let $c_{3\text{D}}$ denote the cylinder's centre point, and e^1 and e^2 its endpoints, i.e., the points of intersection between the cylindrical axis and the planes (see Figure 5.1). Then, we define the marker model M^{mod} as a set of spatial points s such that:

$$\text{M}^{\text{mod}} = \begin{cases} s^T A^{\text{cyl}}\, s \leq 0, \\ (s - e^1)^T (e^1 - c_{3\text{D}}) \leq 0, \\ (s - e^2)^T (e^2 - c_{3\text{D}}) \leq 0, \end{cases} \tag{5.1.2}$$

where the matrix A^{cyl} is comprised by the cylinder's shape, and spatial position and orientation matrices that are computed using the cylinder's radius r, and the points $c_{3\text{D}}$, e^1, and e^2. A detailed derivation of A^{cyl} that is given in Appendix C.3.

The intensities of the points that constitute the projection of the marker model are computed in the following way. We assume that the x-rays attenuate equally at any point inside the marker model. We assume that $\mu(p) = 1$ for the points inside M^{mod} and $\mu(p) = 0$ outside of it. Therefore, we compute an intensity at some point on the isocentric plane, p^{iso}, as the length of the part of the ray from p^{src} to p^{iso} that falls inside the marker model. As shown in Figure 5.1, this is the distance between s^{in} and s^{out}, i.e., the points of intersection between M^{mod} and the ray. Hence, the

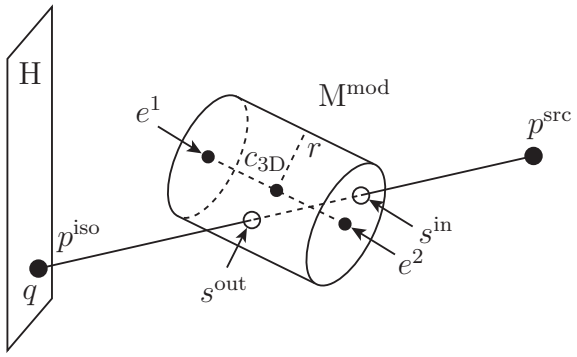


Figure 5.1: Raytracing for the cylindrical model.

projection of one marker model M_k^{mod} results in the following intensity in point q in image $\mathcal{W}_k$:

$$\mathcal{W}_k(q) = \|s^{in} - s^{out}\|, \tag{5.1.3}$$

where $q \in \mathbb{R}^2$ is the point whose coordinates are expressed in the UV coordinate system, and which geometrically corresponds to p^{iso}, expressed in the XYZ coordinates. Under the assumption that the imaging system is in the initial position, the relationship between q and p^{iso} is $q = (-p_z^{iso}, -p_y^{iso})^T$, as defined in Equation (B.5.3) and the accompanying explanations. Finally, the projection images of all marker models $\mathcal{W}_k$ are combined into the MM DRR image as:

$$\mathcal{W} = \sum_{k=1}^{n} \mathcal{W}_k \tag{5.1.4}$$

with the subsequent scaling of the intensities in $\mathcal{W}$ to the interval $[0, 1]$.

Figure 5.2 demonstrates an example of a cylindrical model of radius $r = 0.8$, height $h = 5$ that is rotated at 45° around Y, 45° around X, with the centre at $c_{3D} = (5, 0, 0)^T$, i.e., it is oriented to be out of YZ plane. The centre of projection $(0, 0, 0)^T$ is marked by the *star*. Figure 5.3 shows its integral projection on the plane H computed by the MM DRR method.

▲ δ_a, **Amplification of the marker model size**

To demonstrate the method's potential, Figure 5.3 shows the template image generated from the continuous marker model on a fine grid with the pixel size $\delta_{px} = 0.02$ mm, where the intensity values were computed for the rays going from the source to the grid points, i.e., pixel centres. However, the pixel size in our fluoroscopic images is larger, namely 0.2604 mm at the isocentric plane. Taking into consideration the small marker size, 3×0.8 mm, it may be necessary to shoot multiple rays to cover each pixel sufficiently, where the pixel's intensity will be the mean (or sum)

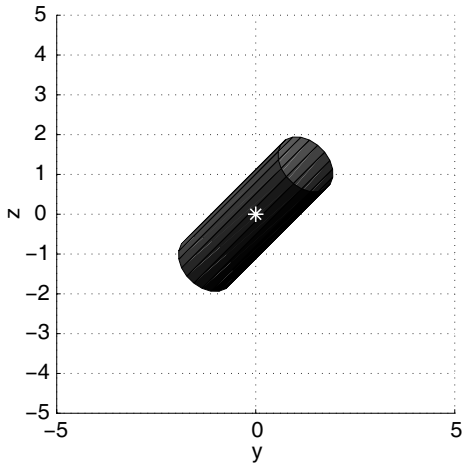


Figure 5.2: Cylindrical model.

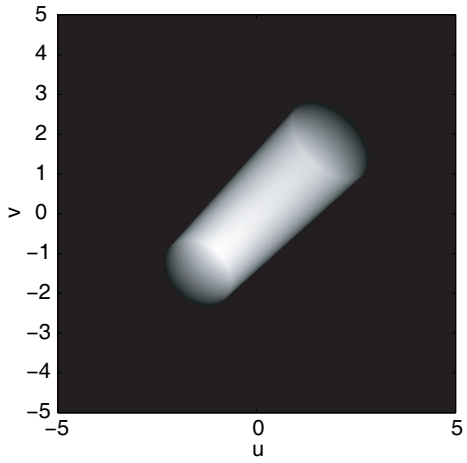


Figure 5.3: Template image.

of the intensities in all points within the pixel. Otherwise, if only the central point within each pixel is used, the marker model's projection may appear smaller than it should be. This effect can be observed in Figure 5.4, where the centres of the pixels marked by *unfilled circles* will produce $\mathcal{W}(\check{q}) = 0$, although these pixels should have some intensity as part of their area is within the marker model projection edge (*solid line*). As the increase in the number of rays inevitably increases the need for computational resources, we took an alternative approach: Enlarging the marker model in 3D before the template image generation requires virtually no additional computation and produces desirable results. In Figure 5.4, the enlarged model and its corresponding projection are shown using *dashed lines*. If the model in 3D is increased by an amount that corresponds to half the fluoroscopic pixel size in 2D, the centres of the pixels whose area is partially covered by the projection of the original model will be included inside the projection of the enlarged model. The enlargement is done by extending the model's radius r by δ_{r} and moving its endpoints, e^1 and e^2, away from its centre $c_{3\mathrm{D}}$ by δ_{h}:

$$\begin{aligned} r_{\mathrm{new}} &= r + \delta_{\mathrm{r}}, \\ e^{1,2}_{\mathrm{new}} &= c_{3\mathrm{D}} + (e^{1,2} - c_{3\mathrm{D}}) \, \frac{\|e^{1,2} - c_{3\mathrm{D}}\| + \delta_{\mathrm{h}}}{\|e^{1,2} - c_{3\mathrm{D}}\|}. \end{aligned} \tag{5.1.5}$$

Ideally, $\delta_{\mathrm{r}} = \delta_{\mathrm{h}} = \delta_{\mathrm{px}}/(2\kappa)$, where κ is the 3D to 2D scaling factor for the size of the object that is located in the proximity to the point $c_{3\mathrm{D}}$ defined in Equation (B.5.5). However, the generated template will have to be registered to the MEF-image. Taking into account that application of the MEF to fluoroscopic images tends to make the markers slightly more rounded, we set $\delta_{\mathrm{h}} = 0$, i.e., the enlargement is only performed for the marker model radius. ■

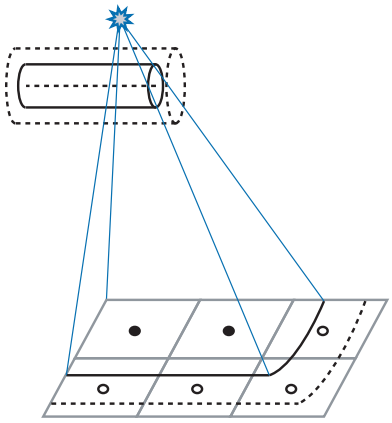


Figure 5.4: Enlargement of the marker model. The real-sized marker model and its projection are shown in *solid lines* and their enlarged counterparts in *dashed lines*. Enlargement of the model ensures that non-zero intensity is assigned to the pixels that contain marker model edge and whose centres are shown by *unfilled circles*.

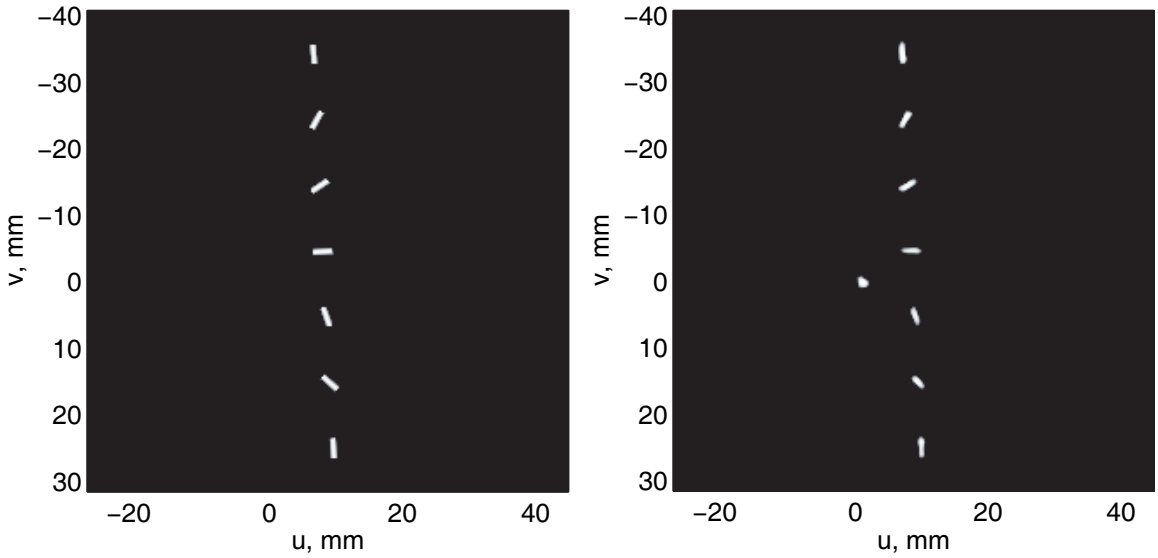


Figure 5.5: Marker model DRR for of a CBCT image of the RSVP (*left*) and the corresponding MEF-image (*right*).

Figure 5.5 shows a sample MM DRR generated from the marker models computed for a CBCT image of the RSVP with 0.5 mm voxels (*left*) and a corresponding MEF-image (*right*), to which the registration will be later performed. The implementation details are provided in Appendix D.3 and summarized in Algorithms D.10–D.12.

5.1.2 Voxel Splat DRR

In this section, we present a voxel splat DRR (VS DRR) method that can be used when the orientation of the marker models can not be computed reliably from the CBCT images, such as for some CBCTs with 1 mm voxels or larger. It is also applicable to irregularly shaped fiducial markers, such as long flexible coils [7]. As opposed to the MM DRR, in which the intensity values at each grid point are computed by raytracing through the marker model, the proposed method aims to form a template image by projecting each CBCT voxel onto the isocentric plane. As can be seen in Figure 5.6, a simple binary image created by projecting the voxel centres onto the pixel grid with the nearest neighbour interpolation to the closest pixel centre can only provide acceptable image quality if the fluoroscopic pixels and the CBCT voxels are of similar sizes, such as $\delta_{\text{px}} = 0.2604$ mm and $\delta_{\text{vx}} = 0.25$ mm (*top figure*). For $\delta_{\text{vx}} > 0.25$ mm (*middle and bottom figures*), the binary images do not produce contiguous projections of the 3D markers, and are not suitable for subsequent image registration to the MEF-image. Therefore, the voxel size should be taken into account.

The proposed VS DRR method is derived from splat rendering techniques [232, 153, 250, 28]. Westover [232] presents a method for image generation from 3D discrete data sets for orthographic views, which can be summarized as spreading a single voxel's intensity onto multiple pixels that lie within the extent of the voxel's projection. Hence, in the generated image, every pixel can contain a sum of contributions from multiple voxels. The method exploits the idea that continuous volumetric data can be obtained by convolving the 3D samples (voxel intensities) with some reconstruction kernel function. Then, the contribution from a voxel to a pixel can be computed by multiplying the voxel's intensity by its *footprint*, which is essentially an integration through the 3D kernel along the viewing direction originating from the pixel centre. In orthographic projection, each voxel will have the same footprint for a particular view. Hence, to omit repeated computation, a footprint table is built once on a fine grid at a preprocessing step, and then the footprint values are interpolated from it for every pixel. The integration along the viewing axis can be substituted by using a 2D cone, Gaussian, sinc or bilinear functions [232]. Birkfellner *et al.* [28] proposed a perspective volume rendering of x-ray images from the CT called *wobbled splatting*. In a perspective projection, each voxel's footprint has to be computed separately, and the

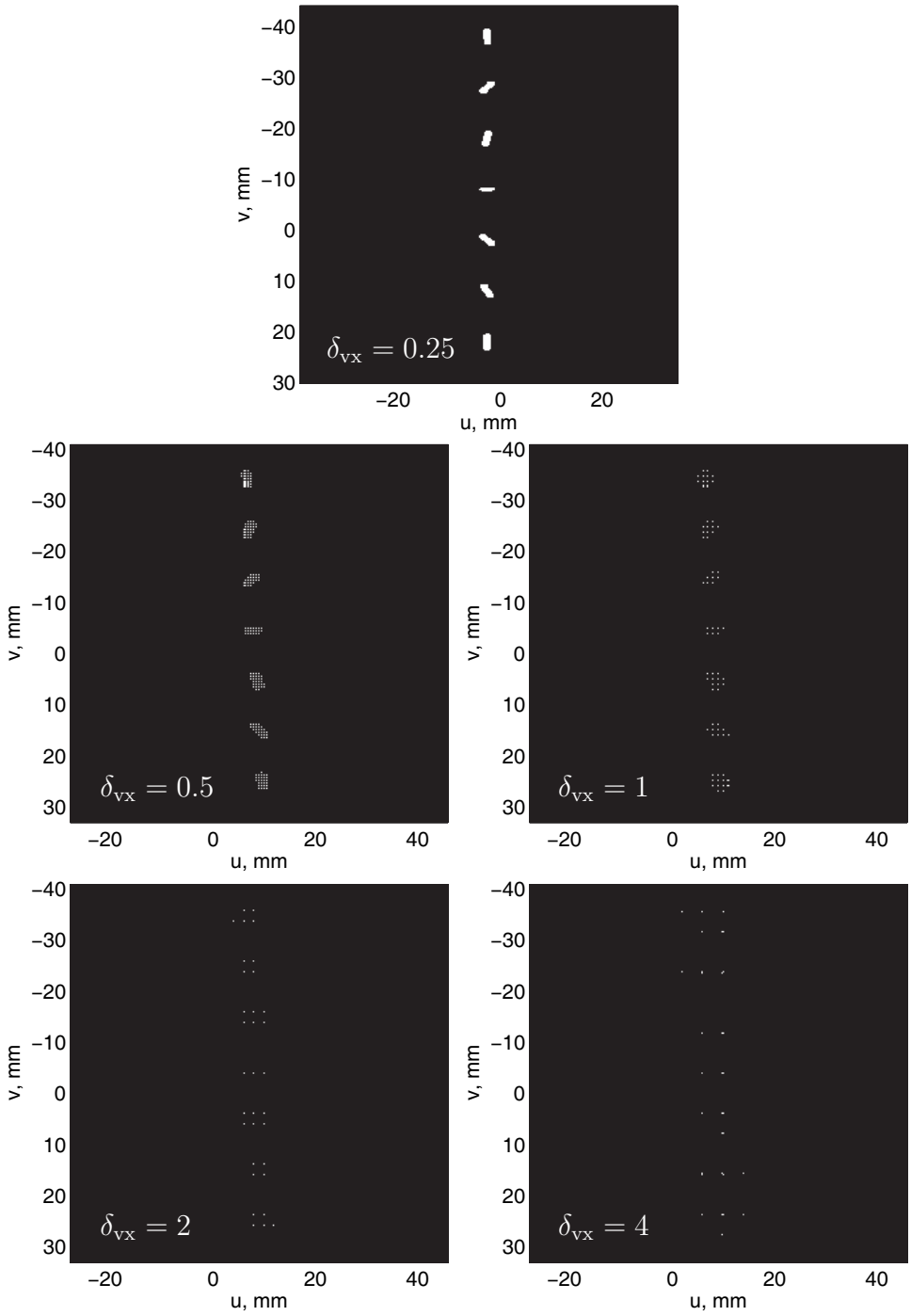


Figure 5.6: Binary images created by projecting the voxel centres onto the pixel grid and using the nearest neighbour interpolation may only demonstrate acceptable quality if the voxels and pixels are of similar sizes (*top figure*) but are of little value if the voxels are larger (*middle and bottom figures*).

computational burden is eased by using only the voxels with intensities above a certain threshold for DRR generation. This is particularly suitable for applications focused on analysis of the bony structures. In [28], a truncated Gaussian kernel is used. The authors argue that the use of a large kernel is too expensive, and propose to overcome aliasing artifacts in the projection images by stochastic motion of parameters in the splatting process, in particular, Gaussian motion of voxels and the viewpoint.

Similar to Westover [232], we use a 2D Gaussian function to model the voxel's integral projection in 2D:

$$\mathcal{G}(q,\sigma) = \exp\left(-\frac{\|q\|^2}{2\sigma^2}\right). \tag{5.1.6}$$

In order to extend the method to perspective projection, we have to take into account the fact that the 3D positions of the voxels (in the relation to the x-ray source and the isocentric plane) influence the sizes of projections of these voxels in 2D. Hence, we modify the method by using the Gaussians of different scales, σ_i, for different voxels. The method remains computationally efficient as a very limited number of voxels are used, namely those belonging to the 3D marker sets segmented as explained in Chapter 3. Unlike in [232], no preliminary footprint lookup table is computed, and no interpolation is required.

Let M_k be the k^{th} 3D marker set, and $|\mathrm{M}_k|$ be the number of points in this set. For every voxel centered at $\check{p}^i \in \mathrm{M}_k$, we put a Gaussian "splat" in the 2D point where the voxel's centre projects, $\mathcal{P}(\check{p}^i)$, where $\mathcal{P}$ is the 3D to 2D projection operator defined in Equation (B.5.4). Hence, a projection of one marker set is defined as:

$$\mathcal{W}_k(q) = \sum_{i=1}^{|\mathrm{M}_k|} w_i\, \mathcal{G}(q - \mathcal{P}(\check{p}^i), \sigma_i), \tag{5.1.7}$$

where the weights w_i are the the corresponding voxel intensities $\mathtt{V}[\check{p}^i]$ scaled to the interval $[0.1, 1]$ as described in Section 3.2. The choice of the Gaussian scales σ_i is explained below. Similar to the MM DRR, the VS DRR is computed by summing the projections of all sets (see Equation (5.1.4)), and rescaling the resulting image to the interval $[0, 1]$.

▲ σ_i, **Gaussian scales of voxel splats**

Appropriate values for σ_i depend on the CBCT voxel size, δ_{vx}, and the positions $\check{p}^i \in \Gamma^3$ of the corresponding voxels. The idea behind the choice of σ_i is that if the marker set points belong to the neighbouring voxels, the sum of the corresponding Gaussians should also form a contiguous high-intensity feature in the 2D template image. Let $\delta_i = \kappa\, \delta_{\mathrm{vx}}$ be the size of the voxel projection to 2D, where κ is the 3D

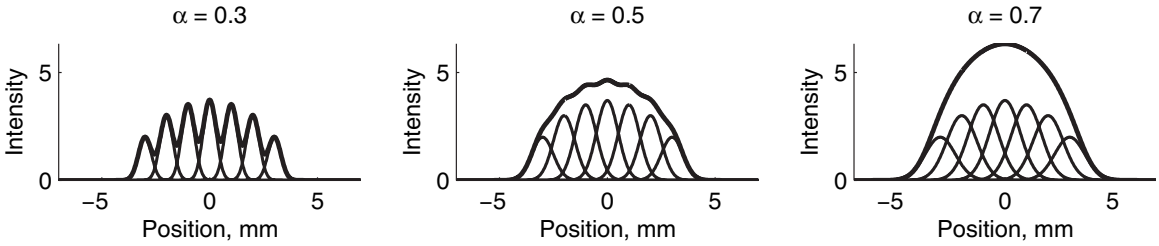


Figure 5.7: The scales σ_i of the voxel splat Gaussians (*thin lines*) affect the smoothness of their sum (*bold lines*). In general, the smoothness increases for larger values of α, as demonstrated from *left* to *right* for $\alpha = 0.3$, 0.5, and 0.7. We select $\alpha = 0.5$.

to 2D scaling factor for the size of the object that is located in the proximity to the point $\check{p}^i$ defined in Equation (B.5.5). Then, the Gaussian scales are computed as:

$$\sigma_i = \alpha\, \delta_i, \tag{5.1.8}$$

where α is a positive scaling factor. Figure 5.7 demonstrates a 1D example of the Gaussians of different amplitudes w_i separated by $\delta_i = 1$ mm with $\alpha = 0.3$, 0.5, and 0.7. As α increases, the individual Gaussians (*thin lines*) become wider, and their sum becomes smoother (*bold line*), which is desirable for image registration. However, it also makes a cluster of high-intensity points larger, which may increase geometric uncertainty when registering to the smaller markers found in the MEF-image. Figure 5.8 shows the template images generated from the marker sets segmented in the CBCTs of the RSVP with 0.5 (*left column*) and 1 mm voxels (*right column*) and different values of α. We conclude that $\alpha = 0.5$ is a good tradeoff for the proposed application. ■

Figure 5.9 shows a sample VS DRR image generated from the marker sets found in a Quasar CBCT image with 0.25 mm voxels (*top*) and RSVP CBCT images with 0.5, 1, 2, and 4 mm voxels (*middle and bottom*). The implementation details are provided in Appendix D.3, and summarized in Algorithm D.13. General considerations and the examples in Figure 5.9 suggest that while the method works quite well for high to medium CBCT resolutions, there is little benefit of using it when the size of the marker is similar to the size of the voxel, such as 2 or 4 mm. That observation led us to the development of the marker splat DRR generation method presented in the next section.

5.1.3 Marker Splat DRR

The marker splat DRR (MS DRR) method is simple and computationally efficient (sample computational times for all three image generation methods are provided in

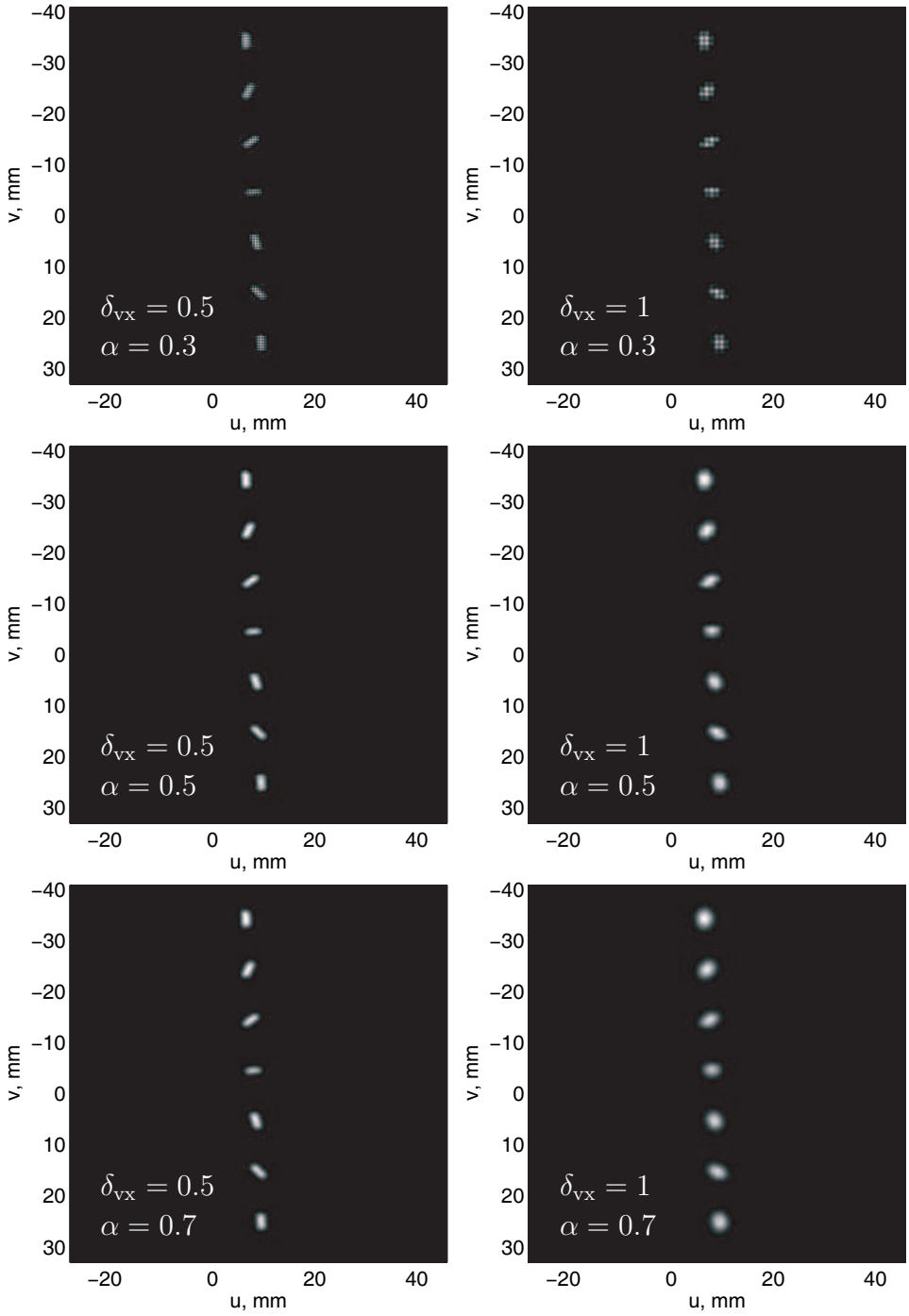


Figure 5.8: Voxel splat DRRs generated for different values of α from the marker sets segmented in the RSVP CBCTs with 0.5 (*left column*) and 1 mm voxels (*right column*). The value of $\alpha = 0.5$ allows to generate contiguous features (compare to $\alpha = 0.3$) but does not oversmooth the markers (such as with $\alpha = 0.7$).

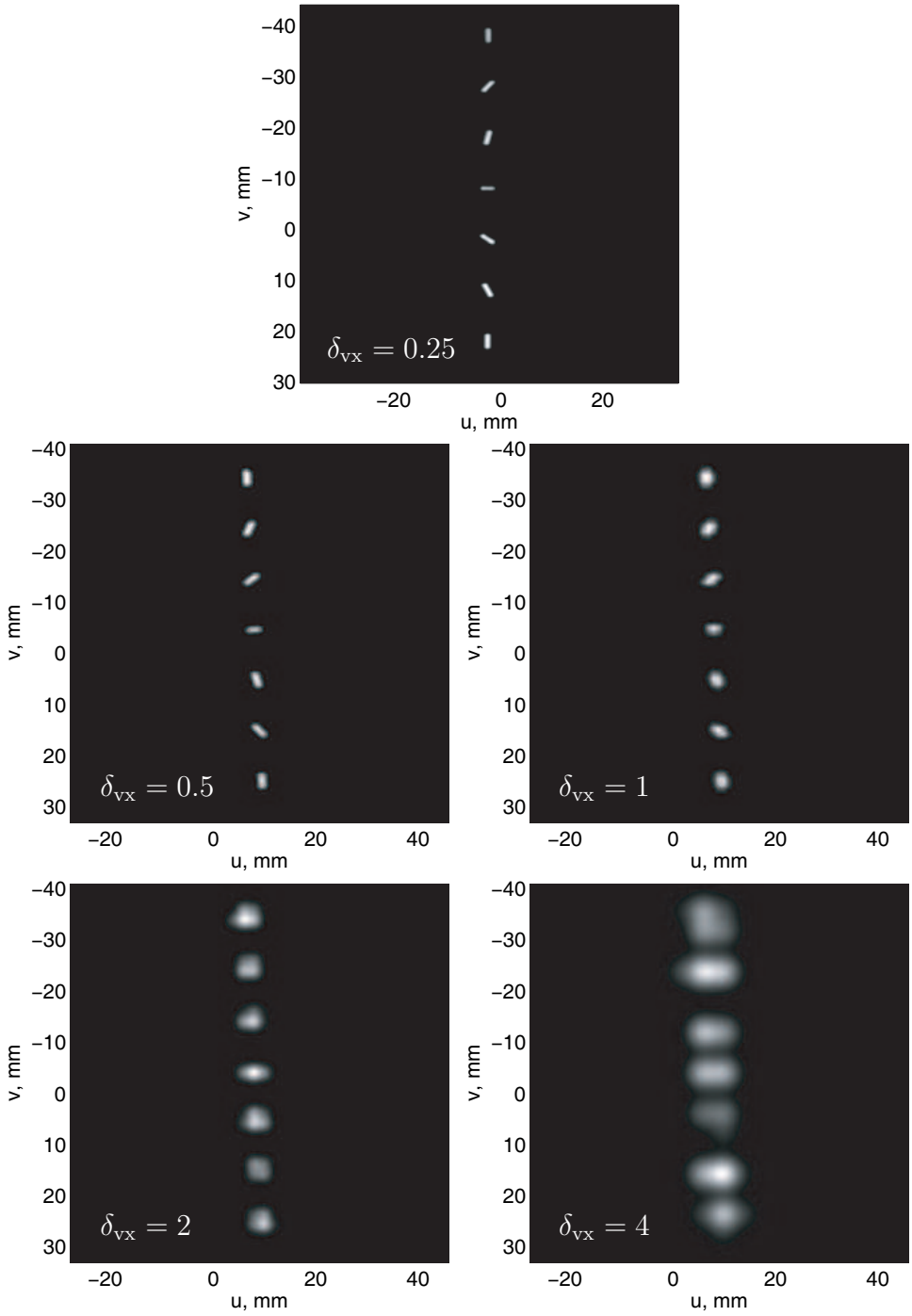


Figure 5.9: Voxel splat DRRs generated from the marker sets found in a Quasar CBCT with 0.25 mm voxels (*top*) and RSVP CBCTs with 0.5, 1, 2, and 4 mm voxels (*middle and bottom*).

Section 7.3.4). It does not take marker orientation into account, only the position, and can be used for CBCT images of any resolution, including images with large voxels. This method can be viewed as a simplified version of the VS DRR described in Section 5.1.2, in which, rather than having the 2D Gaussians for all voxels belonging to the marker sets, one Gaussian (of a larger scale) is used for each marker:

$$\mathcal{W}_k(q) = \mathcal{G}(q - \mathcal{P}(c_{3D}), \sigma), \tag{5.1.9}$$

where c_{3D} is the centre of the k^{th} marker, and the choice of σ is explained below. The MS DRR is formed by summing $\mathcal{W}_k$ (Equation (5.1.4)) and rescaling the resulting intensities to the interval $[0, 1]$. Unlike voxels, which possess various intensities, all markers are treated equally, and therefore the Gaussians in (5.1.9) are not weighted.

▲ σ, **Gaussian scale of marker splats**

In the process of registration between the generated template and the MEF-image, the distance measure function compares the intensities in the images directly or constructs some measures based on the intensities (Section 5.2). Due to this fact, it is desirable that the scales of the marker splat Gaussians are selected so that the clusters of high intensity points created by the Gaussians are of comparable sizes to the markers in the MEF-image, such as schematically shown in Figure 5.10 (*centre*). To estimate an appropriate value for σ, we performed a least-squares fit between the Gaussian (*dashed line* in Figure 5.11) and an idealized marker lengthwise profile defined as 1 for points $-1.5 \leq q \leq 1.5$ mm and 0 otherwise (*solid line*). The optimal value found was 1.1692 mm. Therefore, we select $\hat{\sigma} = 1$ mm. Then, $\sigma = \kappa\hat{\sigma}$, where κ is the 3D to 2D scaling factor for the size of the object that is located in the proximity to the point c_{3D} defined in Equation (B.5.5). Therefore, for different marker splats their scales σ depend on the 3D positions of the corresponding markers. ■

Figure 5.12 demonstrates an MS DRR image generated from a CBCT image of the RSVP with 0.5 mm voxels (*left*) and a corresponding MEF-image (*right*). Note that the size of the high-intensity clusters generated by the Gaussians does not depend on the size of the CBCT voxels, thus making the MS DRR method applicable to the CBCTs with large voxels. Implementation details are provided in Appendix D.3, and summarized in Algorithm D.14.

Summary

To conclude, any of the three methods proposed above can be used for marker specific template image generation. As can be seen in Section 7.3.4, the success rates of image registration with the use of the MM, VS, and MS DRR methods are quite similar on our testing data sets: The MS DRR slightly outperforms the other two approaches by

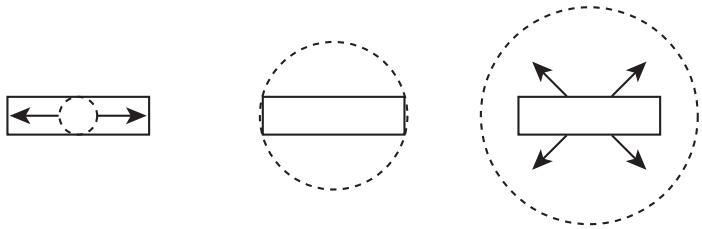


Figure 5.10: The selection of an inappropriate scale for the marker splat Gaussians can increase geometric uncertainty (*left and right*) in the process of registration of the generated template image (*dashed circle*) to the MEF-image that contains the markers (*solid rectangle*). It is desirable that the high-intensity clusters created by the Gaussian are of a size comparable to that of the markers in the MEF-image (*centre*).

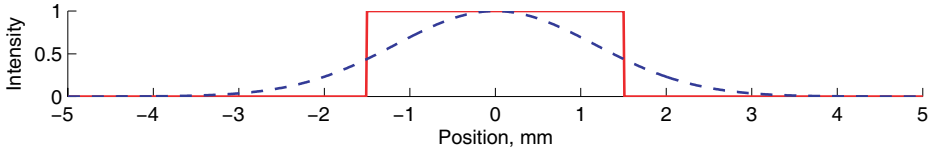


Figure 5.11: The scale of the marker splat Gaussian is selected based on the best least-squares fit between an idealized marker intensity lengthwise profile (*solid line*) and a Gaussian (*dashed line*).

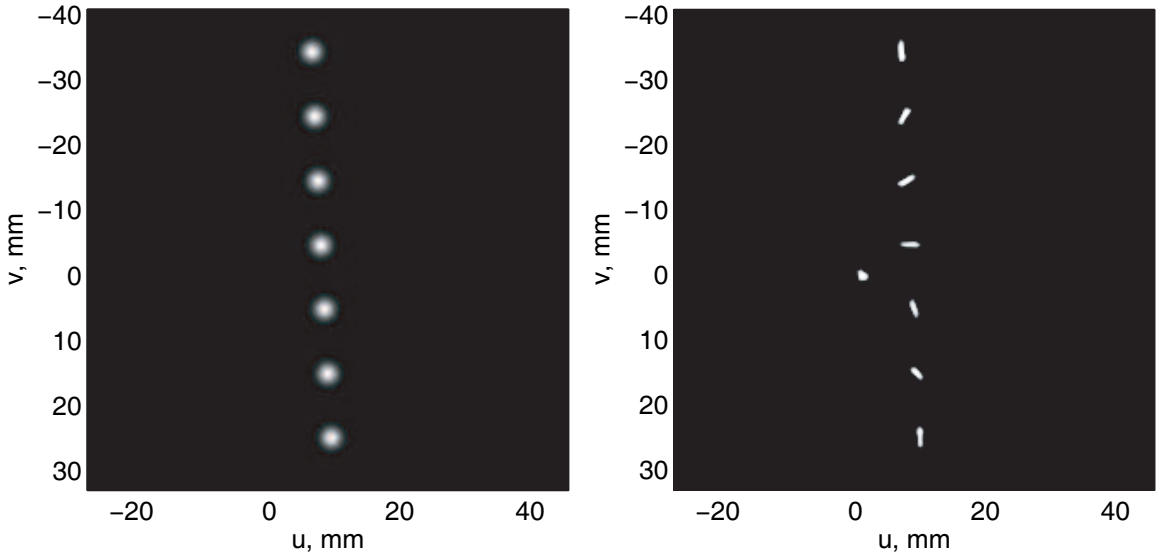


Figure 5.12: Marker splat DRR generated from a CBCT image of the RSVP (*left*) and a corresponding MEF-image (*right*).

successfully registering several more images out of the hundreds of images undergoing registration.

The ability to generate the markers in the template image that closely resemble those in the MEF-image, such as produced by the VS and MM DRR methods, may have the potential of more precise localization compared to the use of the MS DRR method. However, on our testing data sets, the accuracy of all methods was very similar (see Section 7.3.2). In addition, as can be seen from the detailed analysis in Section 7.3.4, the localization accuracy can drop in the presence of streak artifacts and errors in estimation of marker orientations in 3D. Therefore, to process the images similar to those in our current testing data sets, we select the MS DRR method as the most reliable and efficient in our Matlab implementation (the computational cost is discussed in Section 7.3.4), while the VS and MM DRR methods can successfully be used for the CBCT images of better quality. For other applications, where the markers are of an unknown shape or size, in particular, long flexible coils or spiral stents, the VS DRR is likely to be the most appropriate method.

5.2 Distance Measures

Before presenting the details of the proposed 2D-3D image registration problem, we would like to discuss the selection of a distance measure function, $\mathcal{D}$, introduced in (5.0.1), the minimization of which is referred to as step $\langle\langle \mathbf{L}_{3B} \rangle\rangle$ in Flowchart 5.1.

The function $\mathcal{D}$ aims to quantify the difference between the images. Depending on the assumptions about the images, such as whether they belong to the same or different modalities, whether there is a linear dependance between the grayscale values in the images, whether the edges of the structures may provide a reliable base for registration, and the nature of the noise in the images, different distance measures can be used. References [83, 84, 239, 172, 203, 241] contain definitions, discussions and evaluations of the distance measures typically used in medical image registration. For the convenience of the reader, we present a concise review of commonly used functions in Appendix C.4. In this section, we compare these functions in application to our image registration problem and select the most appropriate one. In spite of abundant literature on distance measures, our selection has to be based on experiments with our own images, as both the generated template and the MEF-image are created by the novel processing techniques.

The selection of the distance measure is governed by the necessity to be able to achieve desirable registration accuracy in reasonable time. In general, this requires a function that can produce monotonically decreasing, or at least non-increasing, values towards the desired solution. A common practice to evaluate different distance

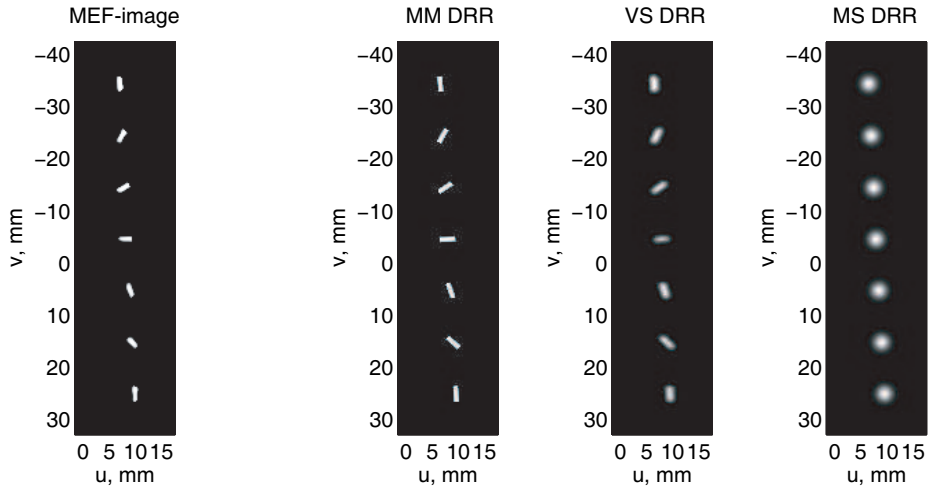


Figure 5.13: The MEF-image of the RSVP and three template images generated from the 3D markers: marker model (MM), voxel splat (VS) and marker splat (MS) DRRs (*from left to right*).

measure functions is to conduct experiments with the typical images for a given registration task, e.g., [203, 83, 172, 241]. Likewise, we select an appropriate function based on logical arguments originating from our specific registration problem, and evaluate the application of the distance measures defined in Appendix C.4 to our typical images. We also compare computational time of our implementations.

In our experiment, we compute 2D profiles of the distance measure functions between the MEF-image of the RSVP and the three generated template images: marker model (MM), voxel splat (VS), and marker splat (MS) DRRs, described in Sections 5.1.1, 5.1.2, and 5.1.3, respectively. The images are shown in Figure 5.13. The distance measure profiles presented in Figures 5.14 and 5.15 were computed by extending the background of each of the the template images, shifting the MEF-image along the U and V directions on a pixel grid,[2] and computing a distance function value for each shift. The distributions of the computational times are shown in Figure 5.16. We observed the following:

- **Intensity differences:** Sum of squared differences, $\mathcal{D}_{\text{SSD}}$ (C.4.1), and sum of absolute differences, $\mathcal{D}_{\text{SAD}}$ (C.4.2). These measures are attractive due to their simple implementation and small computational times. In their conventional definitions [83], the sums of the squared and absolute differences are divided by the total number of pixels in the registration ROI, Ω_{LOC}. This is to account

[2]Shifts are multiples of the pixel size and no interpolation is required. For each image, the distance function was evaluated at 53,361 positions.

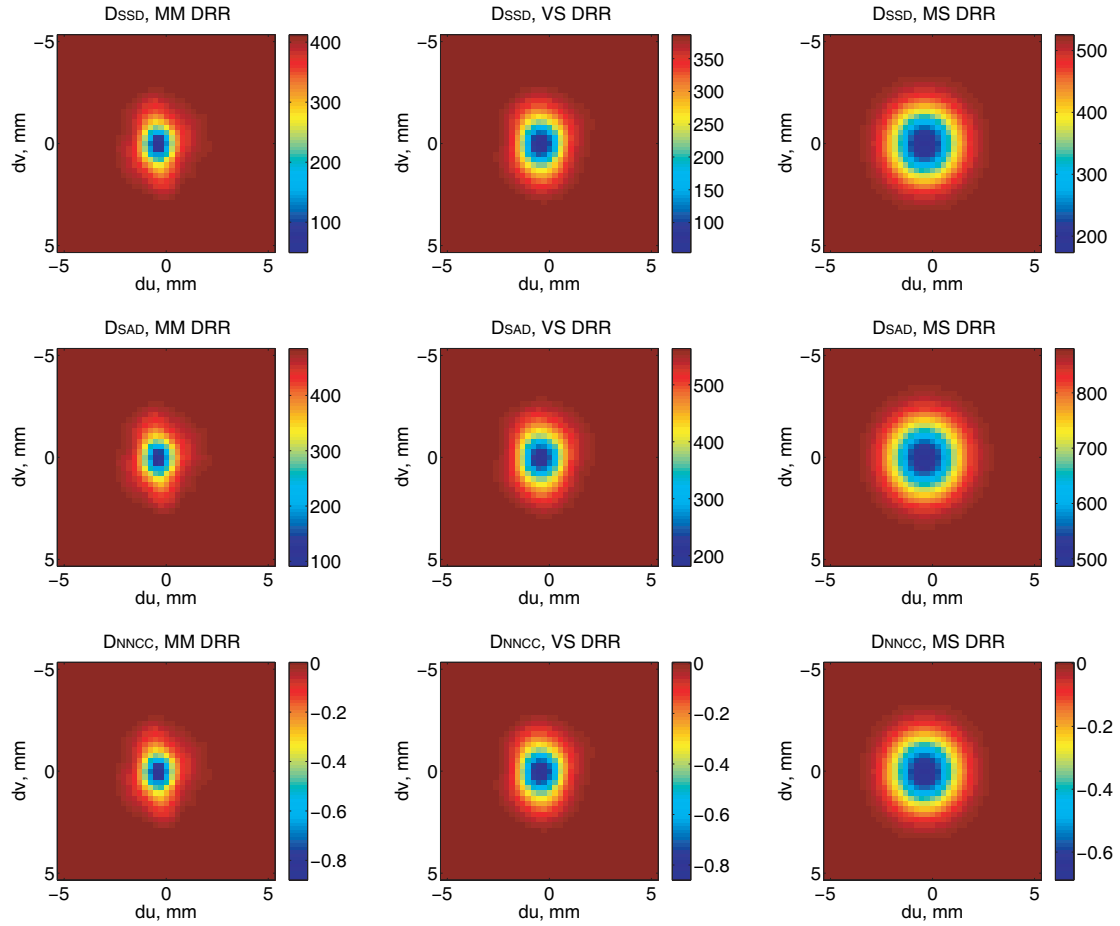


Figure 5.14: $\mathcal{D}_{\text{SSD}}$ (1^{st} *row*), $\mathcal{D}_{\text{SAD}}$ (2^{nd} *row*), and $\mathcal{D}_{\text{NNCC}}$ (3^{rd} *row*) for MM (*left column*), VS (*middle column*), and MS (*right column*) DRRs.

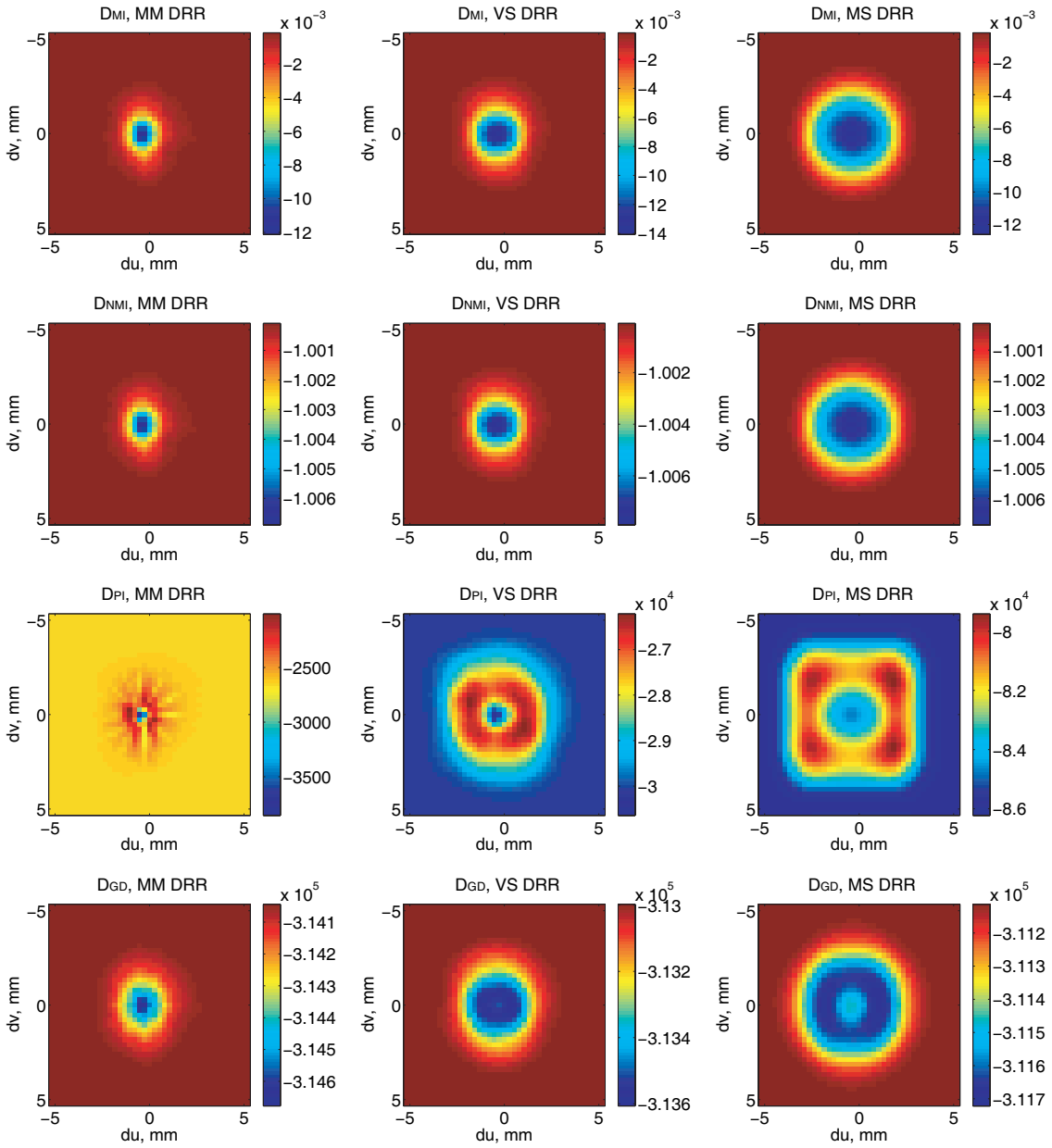


Figure 5.15: $\mathcal{D}_{\mathrm{MI}}$ (1^{st} *row*), $\mathcal{D}_{\mathrm{NMI}}$ (2^{nd} *row*), $\mathcal{D}_{\mathrm{PI}}$ (3^{rd} *row*), and $\mathcal{D}_{\mathrm{GD}}$ (4^{th} *row*) for MM (*left column*), VS (*middle column*), and MS (*right column*) DRRs.

for possible changes in the overlap between the two images, i.e., variations in the size of Ω_{LOC}. This is a reasonable approach for images where the most of the pixels in Ω_{LOC} contain information, i.e., have grayscale values different from zero or other value denoting an empty background. However, in our case the markers occupy a very small part of Ω_{LOC}. Hence, the division by the total number of pixels will result in very small function values that change very slowly, i.e., in an almost flat function shape. This may complicate the design of an optimization method. In particular, for the iterative optimization scheme that we use (Section 5.3.3), it will be challenging to decide on the stopping criteria, such as the minimal change in the function value for which the optimization search should continue. In our case, it should probably be very small to achieve a reasonable registration accuracy. However, having it too small may result in the increase of computational time.

Another possibility is to divide by the number of pixels that contain useful information, i.e., by the estimated number of pixels that are occupied by the markers. However, the extension of this approach to multiresolution or reduced resolution schemes (explained in Section 5.3.3) is not straightforward.

Finally, the scaling can be dropped altogether, such as in the results shown in Figure 5.14 (1^{st} *and* 2^{nd} *row*). Although both $\mathcal{D}_{\text{SSD}}$ and $\mathcal{D}_{\text{SAD}}$ provide reasonable distance measure function shape, the main drawback is that the range of values taken by the distance measures depends on the pixel intensities in the images and on the size of Ω_{LOC}.

- **Correlation:** The negation of normalized correlation coefficient, $\mathcal{D}_{\text{NNCC}}$ (C.4.3). As reviewed in Appendix C.4, the NCC was successfully used in many applications. Similar to $\mathcal{D}_{\text{SSD}}$ and $\mathcal{D}_{\text{SAD}}$, $\mathcal{D}_{\text{NNCC}}$ has a reasonable function shape shown in Figure 5.14 (3^{rd} *row*). In particular, the peak is not too narrow at its base, and is sufficiently sharp around the minimum point, the properties necessary for a reasonable capture range[3] and accuracy. In addition, the NCC has another important advantage: it is scaled from -1 (inverted images) through 0 (no correlation) to 1 (identical images), which makes it independent of the pixel intensities and the ROI size. However, due to the normalization, the NCC can be quite sensitive in regions of near uniform intensity, and has the potential to artificially increase where a marker does not exist. In our application, the background of the generated images is explicitly set to zero. These areas are easily detected, and this sensitivity is mitigated.

[3]The size of the region around the global optimum of a registration problem, for which the selection of a starting point from this region allows a local optimization method to reach the global optimum. More discussion is provided in Section 5.3.3.

- **Information theoretic techniques:** mutual information, $\mathcal{D}_{\text{MI}}$, and normalized mutual information, $\mathcal{D}_{\text{NMI}}$, defined in Equations (C.4.6) and (C.4.7), respectively. These distance measures are mainly used for multi-modality registration, as they do not assume a linear relationship between the intensity values in the images. Although they successfully register our test images, they are more expensive computationally than all other tested measures, as demonstrated in Figure 5.16. Also, they tend to have a slightly wider almost flat area around the minimum in comparison to $\mathcal{D}_{\text{SSD}}$, $\mathcal{D}_{\text{SAD}}$, and $\mathcal{D}_{\text{NNCC}}$, which can reduce registration accuracy due to premature termination of the optimization algorithm. It can be seen in Figure 5.15 (1^{st} *and* 2^{nd} *rows*) for the case of the MS DRR, and to a less extent for the VS DRR, i.e., when the shapes of the markers in the two images are different.
- **Gradient measures:** Pattern intensity, $\mathcal{D}_{\text{PI}}$, and gradient difference, $\mathcal{D}_{\text{GD}}$, defined in Equations (C.4.9) and (C.4.11), respectively. Essentially, both $\mathcal{D}_{\text{PI}}$ and $\mathcal{D}_{\text{GD}}$ rely on the assumption that the edges of the structures in the images should coincide when the images are aligned. Then, they are likely to perform well for the MM DRR, reasonably for the VS DRR, and perhaps not so well for the MS DRR, as in the last case the shapes of the markers in the MEF- and generated template images are considerably different. This is confirmed by the testing, the results of which are shown in Figure 5.15 (3^{rd} *and* 4^{th} *rows*). Hence, these distance measures are not suitable for our registration task.

Based on the analysis above, we select $\mathcal{D}_{\text{NNCC}}$ as the distance measure for the registration problem of the proposed marker localization approach. It is characterized by a reasonable function shape for all three image generation methods, is scaled to a known range, which makes the selection of optimization stopping criteria straightforward, and is efficient to compute, which is important taking into account the need for multiple evaluations within the iterative optimization procedure.

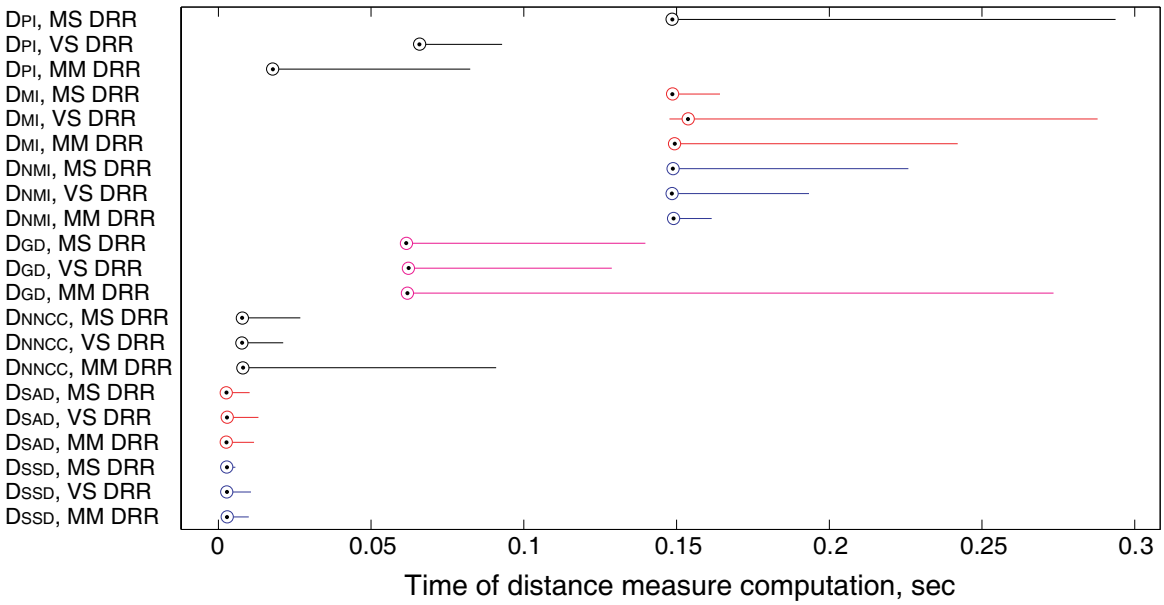


Figure 5.16: Distributions of computational cost for different distance measures computed between the MEF-image and the template images generated with the MS, VS, and MM DRR methods shown in Figure 5.13. The times for $\mathcal{D}_{\text{PI}}$ are significantly different for the MM, VS, and MS DRRs as the number of points for which the computation is performed depends on the size of the markers in the generated image (see Appendix C.4).

5.3 2D-3D Image Registration

Following the discussion of the template image generation methods in Section 5.1 and distance measure functions in Section 5.2, this section puts the pieces together by formulating the 2D-3D image registration problem that is instrumental in the process of marker localization in 2D space, and offering strategies for solving it. It is referred to as step $\langle\langle \mathbf{L}_3 \rangle\rangle$ in Flowchart 5.1. We present details on the geometric transformations in Sections 5.3.1 and 5.3.2, and optimization strategy, including the selection of the starting point for the iterative procedure, in Section 5.3.3.

The top-level formulation for image registration problem was given in Equation (5.0.1). As was explained in Section 2.2, we assume organ rigidity, which allows us to use rigid geometric transformations. Therefore, the search for an optimal transformation $\mathcal{T}^*$ in our case means the search for the vector of optimal transformation parameters a^*, consisting of translations and rotations. Let $\mathcal{R}$ denote the MEF-image, and $\mathcal{W}_a$ the generated template image transformed by $\mathcal{T}(a, \cdot)$. Then, the search for

optimal transformation parameters can be formulated as:

$$\begin{aligned} a^* = \arg\min_{a} \mathcal{D}\left(\mathcal{W}_a, \mathcal{R}\right), \\ \text{s.t. } a \in \Lambda, \end{aligned} \tag{5.3.10}$$

where Λ is a set of *feasible* parameter values, the choice of which will be discussed in Section 5.3.1. The distance measure $\mathcal{D}$, which is referred to as the *objective function*, is computed only over the points in the localization ROI, Ω_{LOC}, found as explained in Section 4.3. In the problem we consider, Λ is compact and $\mathcal{D}(\cdot,\cdot)$ is continuous, and hence by the Weierstrass' theorem [25] a minimizer exists.

5.3.1 Bounds on Optimization Parameters

In the optimization problem (5.3.10), the values of the geometric parameters a are restricted to lie in a feasible set Λ. These constraints are used to ensure that the solution is reasonable: The fiducial markers are, indeed, enclosed within the organ, and can not exhibit an unreasonably large motion within a calmly lying patient.

It is traditional in radiation therapy to describe the patient rigid organ motion in terms of the shifts along and rotations around the three patient axes, namely, the left-right (LR), superior-inferior (SI), and the anterior-posterior (AP) directions. Given the limited number of the predefined treatment positions, it is typical to have the LR, SI, and AP axes of the prone or supine prostate patients aligned to the X, Y, and Z axes, respectively. Therefore, it is natural to express the feasible set Λ as a box, i.e., upper and lower bounds on a: $a^{\min} \leq a \leq a^{\max}$. In this thesis and the proposed localization procedure, it is reasonable to select $a^{\min} = -a^{\max}$, i.e., $|a| \leq a^{\max}$, where $\leq$ is an element-wise comparison.

The values for the elements of $a^{\max}$ can be derived from the maximal observed magnitude of the intra-fraction prostate motion between the CBCT and fluoroscopic image acquisitions or on anatomical knowledge. We select $a^{\max}$ by analyzing multiple studies, which suggest that though the prostate motion is typically described as irregular, it is characterized by the following trends:

- The appropriate dietary protocol, which is currently accepted in some clinics, including the Princess Margaret Hospital where all our prostate patient image sequences were produced, helps to reduce intra-fraction prostate motion [68, 206, 164].
- Prostate motion tends to be smaller in the patient's LR direction (X axis), and larger in AP and SI (Z and Y axes, respectively), as shown in [164, 16, 129] and multiple references in the reviews [68, 116, 38]. This means that the largest prostate shifts should appear in-plane in the lateral fluoroscopic image sequences. It can also be said that the prostate rotates around its apex, and

its motion is mainly parallel to the YZ plane for supine and prone patients. Therefore, the largest rotation in the 3D is also observed in-plane in the lateral images as this is the rotation around the X axis.

- The maximal magnitudes of inter-fraction motion measured in 10 studies reviewed by Langen and Jones [116] cover the ranges of 7–14 mm for the AP, 4–9.9 mm for the SI, 2–9.3 mm for the LR, and 8–20 mm for the 3D displacement. In a number of studies summarized by Byrne [38], it was concluded that the intra-fraction motion tends to be smaller than the inter-fraction motion.

The studies that report on intra-fraction prostate motion differ in the patient preparation protocols, such as dietary restrictions, the positioning particularities, such as the use of immobilization devices and prone or supine patient positioning, the number of patients involved in the study, the length of imaging sessions, the frequency of image acquisitions, and finally, the methods to extract motion data from the images and the formats in which the results are reported. We use analysis of the literature to select such values for the elements of $a^{\max}$ that are likely to include all possible intra-fraction prostate displacements. Therefore, we focus mostly on the maximal motion magnitudes reported.

We analyzed multiple studies that included manual, semi-automatic, and automatic methods of collecting the motion statistics from the images produced using kV x-rays, fluoroscopy, ultrasound, EPID, MRI, and electromagnetic transponders. A number of studies report the intra-fraction motion results in terms of the mean and standard deviation values for the shifts along the patient axes, where all mean and deviation numbers were below 2 mm [129, 108, 163, 174, 8, 68]. Langen *et al.* [117], who used electromagnetic transponders to measure prostate motion, report that the fraction of time during a treatment session for which the prostate was displaced more than 10 mm from its initial position was 0.6%, 0.5%, 0%, and 1.3% for the AP, SI, LR and 3D displacements, respectively. Two studies indicated no intra-fraction motion [190, 224]. The maximal values for the shifts that we discovered in the literature were 23 mm, 24 mm, and 22 mm for the AP, SI, and LR directions, respectively, observed in one of the patients in the study by Kitamura *et al.* [103], and 25 mm in the 3D measured by Kron *et al.* [110]. Most of the other shifts were well within 10 mm, as reported in [88, 16, 103, 163] and referenced by Ghilezan *et al.* [68], with the exception of 16.3 mm shift in the AP direction observed by Rosewall *et al.* [182]. Most studies indicate smaller LR displacements in comparison to the shifts in AP and SI directions.

Concerning the rotations, Aubry *et al.* [16] report the mean ± standard deviations of the rotations around the LR, SI, and AP directions to be $-0.5^\circ \pm 5.8^\circ$, $0.5^\circ \pm 3.8^\circ$, and $0.4^\circ \pm 2.0^\circ$, respectively. As the data on the intra-fraction rotation

is very scarce, we also analyzed the studies regarding inter-fraction rotation. Similar to [16], the three inter-fraction studies [206, 217, 131] also showed the largest rotation around the LR axis. The total mean, the standard deviations of the means of different patients, and the root-mean-square of the standard deviations of different patients are all within 7°.

Based on the literature review presented above, we assume the bounds of 20 mm and 10° for the *in-plane* shifts and rotation (shifts along Y and Z axes, and rotation around X axis in the XYZ coordinate system, or shifts along U and V, and a rotation angle in the UV coordinate system), and 10 mm and 5° for the *out-of-plane* shifts and rotation (shift along X axis, and rotations around Y and Z axes). Although, in general, we expect the prostate motion to be smaller, the literature suggests that large displacements can occasionally occur for some patients [103], and so we believe these large parameter bounds are necessary to tune the system to work reliably with data from different patients.

5.3.2 Geometric Transformation

As mentioned above, we assume organ rigidity between the CBCT and fluoroscopic image acquisitions. Therefore, $\mathcal{T}(a, \cdot)$ is a rigid transformation. Several cases regarding the number of degrees of freedom used in 2D-3D medical image registration are reported in the literature:

- Let t_i denote the shift and γ_i the rotation angle around the I axis. The 3D transformation $\mathcal{T}^{3D}(a, \cdot)$ described in Appendix B.1, where $a = (t_x, t_y, t_z, \gamma_x, \gamma_y, \gamma_z)^T$, models the 3D shifts and rotations of an imaged object in the most accurate way, and is used in a number of studies, e.g., [26, 155, 85].
- However, the imaging system may not be sensitive enough to detect the out-of-plane displacements. Therefore, the use of the 2D transformation $\mathcal{T}^{2D}(a, \cdot)$ described in Appendix B.2, where $a = (t_u, t_v, \gamma)^T$, can be justified, e.g., [7].
- Some approaches intend to determine the 3D shifts and rotations, but argue that minimizing the distance measure in a 6-dimensional search space is prohibitively expensive for some applications [64, 187, 29]. They propose to move a part of the computation to the preprocessing stage [64, 187], or to use geometrical properties of the imaging system to split the problem into the subproblems of lower dimensions [29]. In particular, Fu and Kuduvalli [64] and Sarrut and Clippe [187] propose to compute a number of DRRs where the 3D volume is transformed by a number of the most typical out-of-plane transformations. During the registration in [187], the precomputed DRR that is the most similar to the 2D image defines the out-of-plane transformation parameters, while the in-plane param-

eters are computed from 2D-2D registration. In [64], the 2D transformations from two orthogonal projections are used to estimate the 3D parameters with the use of a limited number of precomputed DRRs. Birkfellner *et al.* [29] observed that the rotation around the viewing axis can be decoupled from the other two rotations and three shifts, and hence, the search space is reduced to five parameters for the first step of the approach. Then, the DRR is generated from the volume transformed by these five optimal parameters. Finally, another registration step between the 2D image and the DRR is carried out with the 2D rotation-only function, which allows to determine the optimal value for the remaining rotation parameter around the viewing axis.

The maximal allowed out-of-plane shifts and rotations that we defined in Section 5.3.1 were 10 mm and 5°, respectively. In our three patient testing data sets, the median 2D distance between the markers in the fluoroscopic images is around 23 mm. If the markers, that are located in the proximity to the isocentric plane, are shifted in 3D by 10 mm out of plane (in the direction towards or away from the x-ray source), the distance between them in the images will change by around 0.23 mm, which is less than a pixel. If they are rotated 5° out of plane, this will translate into a 0.1 mm change in the inter-marker distances, which is less than half of a pixel. The change in the marker size resulting from the maximal out-of-plane shift and rotation is virtually undetectable, being around 0.26 and 0.07 pixels, respectively. Therefore, we conclude that our imaging system may be not sensitive enough to adequately detect an out-of-plane motion.

To investigate the possibility of using the 2D transformation function, we conducted several experiments. As reported in Section 7.3.2, the accuracy of the registration quantified with the fiducial registration error (FRE) is only slightly better for the 3D geometric transformation compared to the 2D: Although the mean ± standard deviation values are quite similar, ranging from 0.09 ± 0.02 mm to 0.11 ± 0.04 mm for the three template image generation methods and 2D or 3D transformations, the maximal FREs for using the 3D function take the range of 0.16–0.23 mm while the 2D transformation is characterized with slightly higher errors of 0.24–0.25 mm.

Another experiment, described in Section 7.3.3, examines the ability of the localization procedure with the 3D geometric transformation to sensibly detect the out-of-plane displacements. The presented results suggest that the imaging system is not sensitive enough to detect these displacements. While the localization accuracy can be just slightly higher, the parameter values pertaining to the out-of-plane displacements are quite meaningless. Finally, the comparison of the computational cost presented in Section 7.3.3 that suggests a 4–6 times speedup for the optimization procedure using the 2D transformation compared to the 3D function ultimately per-

suades us that using the 2D transformation is a reasonable choice for the proposed marker localization procedure.

5.3.3 Optimization

The solution to the minimization problem (5.3.10) can not be found analytically, and we will use an *iterative method* to search for it. In general, iterative methods begin by evaluating the objective function at a starting point a^0, and proceed at each iteration by trying to find another point at which the objective function value is smaller than at the points found at the previous iterations. The optimization routine continues until no further improvement can be found, or, more specifically, until the optimization stopping criteria are satisfied.

Problem (5.3.10) is a *non-convex* constrained optimization problem: Regardless of the distance function properties, a transformation function that includes rotation is not convex, and the optimization search can be trapped in a local minimum. Convexity is discussed in Ben-Tal and Nemirovski [22] and Boyd and Vandenberghe [33]. Multiple *global optimization* strategies are offered in the literature (a review can be found in Pintér [175]). At a high level, most global optimization techniques operate in one of two ways. Either they separate the feasible set into subsets with subsequent comparison of the solutions found locally to select the one with the smallest objective function value, or they replace the original problem by a sequence of subproblems that are easier to solve. In general, these are non-trivial tasks, and in every particular case preliminary knowledge about the problem is likely to be exploited in order to make educated guesses about the separation of the feasible set, the selection of the starting points, or the design of the subproblems/approximations that possess some properties, e.g., convexity. For some problems, it may be possible to eliminate multiple optimization runs and approximations by selecting a good starting point in the proximity to the global minimum.

In the proposed system, the use of the temporal and marker enhancement filtering, as well as the marker specific template image generation approaches, helps to reduce the number of the local minima associated with the *noise* and *discretization* of the images, by forming two images with smooth amplified markers on a flat background before the registration. The choice of the distance measure is also important. For example, consider the smoothness of the $\mathcal{D}_{\text{NNCC}}$ function and the clarity of its minima demonstrated in Figure 5.17 that shows the central horizontal (*left*) and vertical (*right*) 1D profiles of the 2D function profile from Figure 5.14 (*bottom right*, where only the zoom-in into the global minimum is shown). The use of filtering and template generation approaches creates a desired possibility of registering the images by aligning the markers, while mostly excluding anatomical features and background

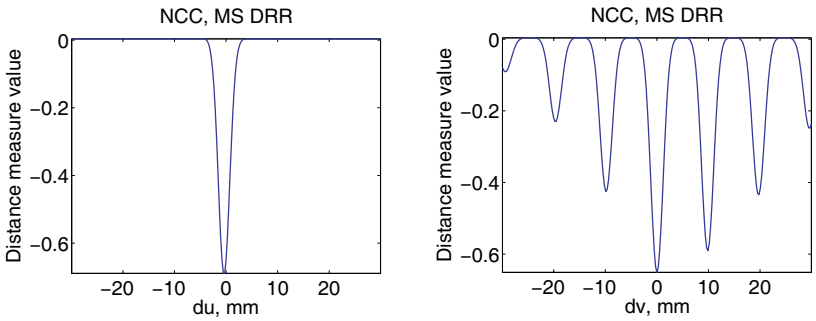


Figure 5.17: A central horizontal (*left*) and vertical (*right*) profiles for $\mathcal{D}_{\text{NNCC}}$ for the MS DRR, the 2D zoom-in into the centre of which is shown in Figure 5.14 (*bottom right*).

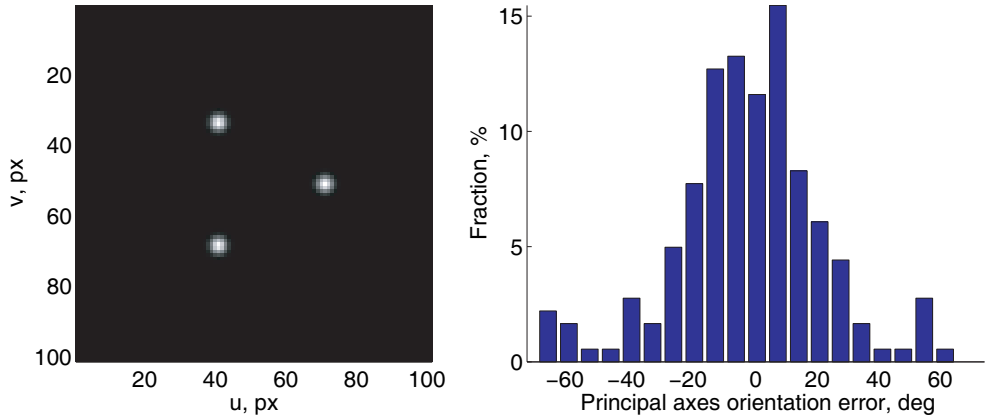


Figure 5.18: For certain high-intensity pixel arrangements, the method of principal axes transformation does not produce desirable results.

variations from consideration.

Unfortunately, this also creates a pitfall: It may be challenging or impossible for an optimization method to find a direction in which the distance measure decreases if the markers in the two images are too far from overlap, i.e., outside of the *capture range* of the distance measure and optimization method. For example, in Figure 5.17 (*left*) the shape of $\mathcal{D}_{\text{NNCC}}$ is flat around the shifts $du = 10$ or 20 mm along the U axis, which means that it may be unlikely for an optimization method to find a decreasing direction if $a^0 = (10, 0)^T$ or $a^0 = (20, 0)^T$. A second issue is the existence of multiple profound local minima, corresponding to matching of the wrong markers in the two images, that can be seen in Figure 5.17 (*right*). Although our phantom images are especially prone to this problem due to the periodicity of the marker placement (see Figure 5.13), this can happen to the patient images, too.

The problems of an insufficiently large capture range and multiple local minima of the distance measure function are common in many medical image registration applications. Several methods to overcome these issues are described in the literature. The simplest one is manual selection of the starting point [135, 72]. As this method is time consuming, and is not suited for automated processing of large number of images, we have excluded it from consideration. The methods we considered include the following:

- **Pre-alignment by principal axes transformation** [149, 145, 9]. In this method, the image is treated as a density function or mass distribution. The centroid of the image points, weighted by their intensity, is used as the origin, and the eigenvectors of the covariance matrix of these weighted points as the axes of the object reference system. Hence, bringing into alignment the centroids and axes of two images of similarly shaped objects yields their approximate registration. However, the method may not be reliable for certain arrangements of the high-intensity points. Figure 5.18 (*left*) shows an example of a generated template image with three equidistant marker splats. When this image is rotated around its centre, i.e., the object reference system is rotated, the principal axes should rotate by the same angle. However, as can be seen in Figure 5.18 (*right*), because of this particular arrangement of high-intensity points and image discretization, the difference between the rotation angle of the object and its principal axes goes beyond $\pm 60°$. Also, the centroid and principal axes can be influenced by the noise and anatomical features that survive filtering. Therefore, this technique can not be used in the unsupervised method that we seek to develop.
- **Multiple optimization runs** [13, 213]. This method operates by selecting a number of starting points, usually either randomly or by sampling the search space uniformly, and runs optimization routine starting from each of these points. Then, the solution that provided the best objective function value is selected. To increase the probability of convergence to the global optimum from at least one of the starting points, the domain has to be properly sampled. This approach can potentially be more computationally expensive than other methods, which is not desirable in our marker localization procedure.
- **Exhaustive search over sampled parameter space** [7, 26, 101] refers to directly evaluating the distance measure function in a number of points from the domain, and selecting the one that provides the smallest function value as either the solution or the starting point for a more sophisticated optimization procedure. Typically, the points are selected by sampling the search space uniformly. Similar to the case with the multiple optimization runs, the domain

should be appropriately sampled to achieve acceptable accuracy or increase the probability of finding a starting point from which an optimization method can converge to the global optimum. This requirement may result in significant computational cost, unless done for a simple geometric transformation and a small number of points. For example, Adamson and Wu [7] performed an exhaustive search with normalized correlation coefficient (NCC) at image pixel resolution, i.e., a 2D shift only with the step size of one pixel in U and V directions.

- **Multiresolution** [155, 212, 20, 37, 150]. Also referred to as hierarchical or multilevel strategies, multiresolution includes a number of methods [119], and is generally described as a "coarse-to-fine" framework. Although varying the complexity of geometric transformation or image registration method can also be referred to as hierarchical approach, in this thesis by multiresolution we mean a class of methods where the hierarchies consist of the levels of changing *data complexity*. In this context, the goal of a multiresolution approach is to smooth out the details in the images. Smoothing out is often combined with a reduction in size. This is done to perform a rapid pre-registration between the images on a coarse scale to reduce the possibility of a local optimum resulting from matching wrong details. Then, the registration result from the coarser level is used as the starting point for the registration at the next finer level, and registration accuracy is expected to improve as the resolution increases.[4] The image at the finest level typically corresponds to the original image. The filtering of the image to obtain the coarse version (perceived as blurring) is often performed by convolving the image with the Gaussian function [155, 212, 20, 37] or by averaging the intensity of the adjacent pixels [150].

After analyzing the possibilities listed above in application to our registration problem, we decided to focus our attention on multiresolution and exhaustive search approaches. We evaluated the following strategies for the selection of a starting point for the optimization procedure:

- **Multiresolution with the Gaussian pyramid starting with a simple projection (MR).** The image at each successive level is formed by applying Gaussian smoothing with an increasing scale and downsampling the image at the previous level [155, 212, 20, 37]. Both the increase in scale and the downsampling are typically done with a factor of 2. The stack of images formed by this approach is referred to as the Gaussian pyramid. The starting point for the optimization at the highest level (most coarse) corresponds to a simple projection of the 3D marker positions to the isocentric plane (in the UV coordi-

[4] This approach is referred to as sequential multiresolution registration. For some applications, it may be more beneficial to perform registration between all levels simultaneously [212].

nate system), i.e., no displacement is assumed and a^0 is a vector of zeroes. The intuition behind choosing the number of levels is that the images have to be blurred until the markers form (i) a contiguous feature that (ii) is large enough to overlap in the two images. For our images, we used six levels.

- **Normalized 2D cross-correlation with multiple templates (NCC-MT).** The method of computing normalized correlation coefficient at the pixel level resolution, such as used by Adamson and Wu [7], is referred to as the *normalized 2D cross-correlation* (NCC). Matlab offers an efficient implementation for this exhaustive search approach in its `normxcorr2` function, which we only perform on the registration ROI. However, due to the fact that rotations are expected, we upgraded the method by creating several template images rotated in-plane. We will refer to this method as the NCC-MT. The number of the templates and their rotation angles are selected based on the geometry of the marker placement. The maximal 2D distance between the UV origin and any of the markers in the fluoroscopic images of the three patient testing data sets is around $d \approx 31$ mm. For the fiducial markers of the size 3×0.8 mm, let us assume that the representative marker dimension is $s = 2$ mm. To efficiently cover the space of possible rotations, we want to choose the rotation angles α for the templates so that there exist two templates, for which the positions of some point, whose distance to the origin is d, are not further apart than s after the rotation is applied. This angle can be computed as $\alpha = 2\arcsin(0.5s/d) \approx 3.7°$. We select $\alpha = 3.5°$, and compute templates for the in-plane rotation angles $\{\beta_i\} = \{-n\alpha, -(n-1)\alpha \ldots, -\alpha, 0°, \alpha, \ldots, n\alpha\}$, where $n \in \mathbb{N}$ and $n\alpha \leq \gamma^{\max}$. In our localization procedure, $\gamma^{\max} = 10°$, therefore, $\{\beta_i\} = \{-7°, -3.5°, 0°, 3.5°, 7°\}$. The rotation angle of the template that achieved the highest NCC value is selected as the rotation angle for the starting point, $\gamma^0 = \beta_i$. As for the shifts for the starting point, t_u^0 and t_v^0, it has to be taken into account that in the NCC-MT approach, the rotation is performed before the shifts, while the geometric transformations defined in this thesis and used in our marker localization and tracking procedures assume otherwise (see Appendix B). Therefore, t_u^0 and t_v^0 have to be computed from the shifts that provided the maximal correlation value and the rotation angle γ^0 as detailed in Appendix D.4.1.
- **NCC-MT on reduced resolution.** In order to reduce the computational burden associated with the NCC-MT method, we tried to reduce resolution of the MEF- and the multiple template images before the NCC is computed. The pixels in the reduced images are computed by averaging 3×3 and 5×5 blocks of pixels in the original images.
- **Multiresolution with the Gaussian pyramid and NCC-MT (MR +**

NCC-MT). Finally, we have also implemented a method based on the Gaussian pyramid, where a smaller number of multiresolution levels is used, namely 3, and the MT-NCC method is applied only at the coarsest level to obtain a starting point for the registration.

All strategies described above are aimed at finding a reasonable starting point a^0, such that an optimization method that seeks a local optimum can be used to solve Problem (5.3.10). We do not consider global optimization approaches. The most frequently reported optimization methods used in medical image registration are the Powell's method [29, 41, 184], Nelder-Mead downhill simplex [7, 42, 101, 213], steepest gradient descent [171, 167, 85, 102], conjugate gradient [104, 135], Quasi-Newton [12, 70], and Levenberg-Marquardt methods [135], the detailed descriptions of which can be found in the literature [45, 179, 104, 135].

To solve our optimization problem, we employ *sequential quadratic programming* (SQP) approach, where a quadratic programming subproblem is solved at each iteration [165]. We used an *active set* implementation of the SQP provided in Matlab `fmincon` function. It demonstrated shorter computational time when applied to our registration problem with similar registration success rates and accuracy in comparison to other available methods, such as the *trust region* and *interior point*. The average optimization run with the active set was about 6 times faster than with the interior point, and 15 times faster compared to the trust region method.

To test the approaches for the selection of the starting point for the optimization procedure described above (MR, NCC-MT on full and two reduced resolutions, MR + NCC-MT), we conducted an experiment on our phantom and patient testing data sets, the details and results of which are described in Section 7.3.5. In the experiment, we compared the marker localization success rates and average running times of the registration procedure in our Matlab implementation. The following characteristics were observed:

- The 6-level MR method was not able to reach a 100% success rate on any of the patient testing data sets. Combined with a significant running time, this method proves inadequate for the proposed marker localization procedure.
- The remaining four methods that used NCC-MT showed 100% success rates on the test images. NCC-MT on full resolution and MR + NCC-MT methods have similar execution times, which are consistently larger than those of the NCC-MT on reduced resolutions, for both 3×3 and 5×5 pixel averaging approaches. The use of reduced resolution results in an implementation that is 3–5 times faster compared to the full resolution.
- The 3×3 and 5×5 pixel averaging methods are characterized with similar running times. Taking into account small size of the markers, 3×0.8 mm,

which in our fluoroscopic images results in about 12×3 pixel measurement, we consider the NCC-MT method with 3×3 pixel averaging to be the most reasonable choice, as the one that achieves efficient computation but does not oversmooth the markers.

5.4 Selection of Temporal Filter Width

Taking into consideration that we have achieved a 100% registration success rate on the testing data sets, as discussed in the previous section, the registration success itself can not guide the choice of the temporal filter width, ℓ, that will be used in the final configuration of our marker localization procedure. Therefore, we will base the decision on general considerations and total running time of our Matlab implementation of the 2D localization procedure. The use of temporal filtering improves the contrast-to-noise ratio (CNR) of the images, and while all images from the testing data sets were registered correctly even without being temporally filtered ($\ell = 1$), we expect that images of other patients may have a lower CNR. Section 7.3.6 reports computational cost of the 2D localization procedure depending on ℓ. The trend is such that running time decreases for the larger values of ℓ. This is due to diminishing computational cost of the MEF processing and optimization, as the process of grouping points in the MEF and image registration are faster when executed on smoother images.

We select $\ell = 5$, as the rate of time reduction becomes smaller around this value. It can be viewed as being large enough to significantly improve the CNR, while being small enough so as not to introduce too much blur for the moving markers: For fluoroscopic image acquisition of 5.5 Hz, the use of $\ell = 5$ allows to create an image from the measurements acquired over an interval of just less than one second.

5.5 Failure Detection

In spite of many studies discussing topics in medical image registration, the literature on failure detection is not abundant, and is mainly devoted to detecting tracking errors, e.g., [196, 199, 214]. Some of these ideas may also be applied to localization failure detection. Sharp *et al.* [196] introduced a set of error indicators that were used to train an artificial neural network classifier to detect tracking errors. The first indicator, a pattern recognition score, was based on the goodness of fit between the marker template and the x-ray image (analogous to a distance measure value in our localization procedure). The second indicator was the distance between rays: The tracking was performed by two x-ray imaging systems, and so a 3D position of a marker

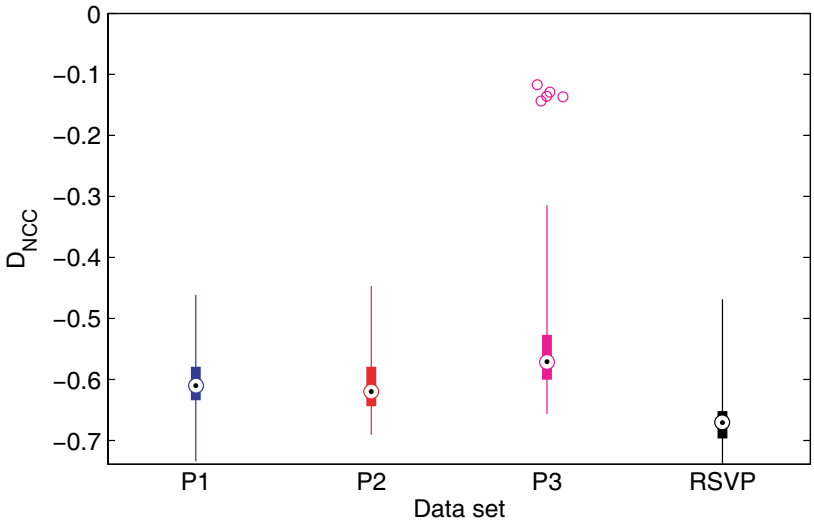


Figure 5.19: Distributions of the objective function values, $\mathcal{D}_{\mathrm{NNCC}}$, for the P1, P2, P3, and RSVP data sets. The outliers (*empty dots*) for the P3 data set correspond to registration failures on extremely noisy fluoroscopic images resulting from imaging system issues, on which a human observer failed to localize the markers as well.

should have been close to the intersection of the rays going from the x-ray source to the 2D markers. These first two indicators were also used by Shirato *et al.* [199]. In addition, the instantaneous velocity and acceleration of the marker were measures of tracking stability over a short period of time. The system included a number of manually set parameters, such as thresholds on error indicators and maximally allowed errors. Tang *et al.* [214] used similar error indicators: an unreasonably small correlation value between the template and the image; a large distance between the marker positions in consecutive frames; and a significant change of the geometric relationship between the markers that were tracked separately in consecutive frames. Each detected marker was viewed as a hypothesis, whose probability to correspond to a real marker was numerically evaluated by several scores dealing with intensity, position and track's correspondence to a breathing waveform. Thresholds on indicators and scores were set empirically.

In the proposed marker localization module, we implemented three procedures for failure detection and prevention. The first one is equivalent to comparing the smallest found value of the distance measure to a manually set threshold similar to other studies [196, 199, 214], and the other two are specific to our methods and implementation:

- **Optimal distance measure value.** In general, a successful registration should

result in a reasonably small objective function value. In our case, there is a hope of expressing the notion of being "reasonably small" numerically, as the objective function, i.e., the distance measure function $\mathcal{D}_{\text{NNCC}}$, is a negative of the normalized correlation coefficient, where the latter ranges from -1 (inverted images) through 0 (no correlation) to 1 (perfect match).

To analyze the distributions of the values of $\mathcal{D}_{\text{NNCC}}$ on our data, we conducted the following experiment: All images from all fluoroscopic sequences of the Patient 1, 2, and 3, and the RSVP, described as the P1, P2, P3, and RSVP data sets in Section 7.3.1, were registered to their corresponding CBCT images, resulting in 1618, 1619, 820, and 423 registrations, respectively. We used the 2D rigid geometric transformation and the MS DRR generation method. The distributions of values of $\mathcal{D}_{\text{NNCC}}$ for each data set are shown in Figure 5.19. In this experiment, the 75^{th} percentiles (*top of the bold bars*) for all data sets fell below the -0.5 value, the largest among the data sets maximum was around -0.3 (*thin bars*), and the registration failures (*empty dots*) were concentrated around the -0.1 value. The registration failures happened in the cases of extremely noisy fluoroscopic frames found in the end of one of the sequences that resulted from technical issues with the imaging system, and for which the successful localization by a human observer was also not possible. Therefore, the optimal objective function value can prove useful for failure detection in our localization problem. Based on the experiment described above, we suggest that a registration producing $\mathcal{D}_{\text{NNCC}} \geq -0.2$ is a candidate for localization failure.

- **Optimization exit flag.** In addition to the optimal parameters and the objective function value, `fmincon` can also return an `exitflag` parameter that describes the reason for stopping of the optimization method. A successful registration corresponds to the situations when the change in the objective function value or the parameters was less than a predefined tolerance, meaning that the point is in the neighbourhood of a local optimum. The termination due to reaching a maximum number of iterations or function evaluations is a candidate for localization failure, and the situation where no feasible point found is definitely a failure.
- **Feasibility of the starting point.** After the starting point a^0 is computed by the NCC-MT method, as described in detail in Section 5.3.3 and Appendix D.4.1, it is verified that all of its elements satisfy $|a_j^0| \leq a_j^{\max}$. The point a^0 can be infeasible due to the following reason: As explained in Section 4.3, the registration ROI within the MEF-image is formed as the minimal rectangular area enclosing all projections of the 3D marker positions that is extended in the U and V directions by a margin equal to the maximally possible 2D marker

displacement. This displacement is computed based on information about the geometric transformation used in registration and the bounds on transformation parameters, $a^{\max}$. While such ROI construction is simple to implement and guarantees the inclusion of all possible 2D positions that the markers can take in the MEF-image as a result of expected organ motion, it also includes the positions that violate the bounds $a^{\max}$. If $|a_j^0| \leq a_j^{\max}$ is not satisfied, the current maximal correlation value in the cross-correlation matrix[5] is removed from consideration, and the search for the maximum and the corresponding shift parameters continues until the inequality is satisfied.

If a failure is suspected or detected, especially in potential real-time applications, it may be overcome by restarting the registration procedure with adjusted parameters. In particular, it may be necessary to reduce the registration ROI if it includes prominent anatomical features that survive filtering; enlarge the ROI, if the underlying assumptions of the marker motion are underestimated; or increase the number of frames used in temporal filtering to improve the CNR of the markers. In our current implementation aimed at post-processing, a warning message is generated.

5.6 Discussion

Although the image registration framework and its components are widely explored in the literature, e.g., [149, 150, 19, 57, 75, 143, 69, 83, 138, 81, 119, 145, 34], the selection of the components and strategies for a particular registration problem may require considerable analysis of the data, the problem, and the algorithms, and this is indeed the case in our application.

According to the clinical protocol, the markers are inserted into the patient's prostate in order to provide positional guidance during the radiation treatment, as they possess significantly better contrast in the images than the soft tissues (Chapter 2). Therefore, in the process of registration between the CBCT and the fluoroscopic images, we are interested to align the markers more than any other features. To achieve this, we designed the marker enhancement filter (MEF, Section 4.4), to preprocess fluoroscopic images. To create a 2D template from the CBCT image that is similar to the MEF-image, where the markers are amplified and anatomical features and background variations are suppressed, we proposed marker specific template image generation approaches in Section 5.1. While all three methods provide similar marker localization accuracy (Section 7.3.2) and success rates (Section 7.3.4), we

[5]Recall that the shift elements of a^0 are computed from the position of the maximum in the cross-correlation matrix. The details are provided in Appendix D.4.1.

have selected the marker splat (MS) DRR method for the proposed localization procedure as the most computationally efficient and better suited for the CBCT images of average or low resolutions that may contain reconstruction artifacts.

We proceeded by selecting an appropriate distance measure among the commonly used functions in application to our testing data sets in Section 5.2. Then, in Section 5.3, the formulation of the registration problem was presented. The bounds on the geometric transformation parameters were determined based on the thorough literature review in Section 5.3.1. Then, the analysis of the geometry of the imaging system and marker placement in Section 5.3.2, supported by the experiments described in Section 7.3.3, suggested that it is more reasonable to use the 2D rigid geometric transformation than the 3D, as the system was insensitive to out-of-plane marker displacements. This choice was also cheaper computationally, as a smaller number of variables were used in the optimization problem.

To perform minimization of the distance measure in the registration problem, we have used the sequential quadratic programming in a standard Matlab implementation of the active set method. It performed similarly in terms of localization accuracy and registration success rates but offered significant computational savings compared to the trust region, interior point, and IPOPT solver. As we decided to use a local optimization technique to solve the non-convex registration problem, we have analyzed the literature and conducted experiments to suggest an appropriate strategy for the selection of the optimization starting point (Section 5.3.3). In terms of capture range and computational cost, the best performance was shown by the normalized 2D cross-correlation method with multiple templates (NCC-MT) performed on reduced resolution. We designed this approach as an upgrade of a standard NCC technique.

The overall performance of the 2D marker localization procedure on our testing data sets was used to select the kernel width of five frames for the Dolph-Chebyshev temporal filter in Section 5.4. Finally, we presented a brief literature review on failure detection and prevention techniques, and designed the measures appropriate for the proposed marker localization problem in Section 5.5.

The results of all experiments evaluating the accuracy and the performance of the 2D marker localization procedure and its components are collected in Section 7.3. Some implementation details are provided in Appendices D.3 and D.4.

Chapter 6

Motion Tracking in 2D Image Sequences

This chapter presents the design of the motion tracking module, which is used to track the markers in the 2D fluoroscopic sequences. It is referred to as block $\langle\langle \mathbf{T} \rangle\rangle$ in Flowchart 2.1 and is detailed in Flowchart 6.1. The literature on image-based fiducial marker tracking with application to radiation therapy can be roughly divided into two approaches: *independent 2D localization* of the markers in each image of the sequence[1] [7, 98, 141, 169, 174, 233, 122], and the use of some *tracking model* that aims to direct and improve tracking by providing a reasonable prediction for the positions of the markers in every new image based on their positions in the preceding images, assumptions on characteristics of motion, and marker placement geometry [195, 214, 186, 140, 142]. Based on the literature, we can separate the process of model-based marker tracking into several steps, such as initialization, prediction, measurement/prediction update, measurement assessment/correction, and tracking failure detection, each of which will be discussed in more detail below. In fact, the first approach (independent 2D localization) can be treated as a simplistic tracking model without the prediction stage.

The most commonly reported realizations of the steps in a model-based tracking system are the following:

- **Initialization**: At this step, the marker positions are defined or a reduction of search space is performed in the first image of a sequence. This can be fulfilled by manually selecting the markers in a 2D image [142, 214, 122], predicting their positions from the CT image [199], or reducing the search to a ROI that

[1] Recall that a fluoroscopic sequence may require the application of some preprocessing techniques, e.g., the temporal filtering, to achieve reliable tracking. Therefore, in this chapter a reference to an "image" means one image of a preprocessed sequence rather than an individual frame of an original fluoroscopic sequence.

can also be either defined manually [98, 233] or from other images, such as CT, CBCT, or DRR [7, 141, 140, 169, 234].

- **Prediction**: The task of this step is to provide a reasonable prediction for the marker positions in the incoming image, or, more generally, at a certain instant of time. For real-time tumour tracking and gating applications, where the time for hardware reconfiguration may be necessary, prediction is crucial. Having a reasonably accurate prediction also allows for more reliable and timely processing, mainly resulting from the narrowing of the parameter search space or a reduction of the ROI. This is beneficial for both real-time and post-processing applications. The most commonly used prediction models are the zeroth-order predictor[2] [142, 141], linear regression or extrapolation [195, 40, 214], artificial neural networks [195], Kalman [195, 186] and particle filters [186]. More details on prediction models will be provided in Section 6.2.
- **Measurement/Prediction update**: At this step, the positions of the markers are found in the current image, with or without prediction. The reported methods include a search for local intensity peaks in the neighbourhood of the predicted positions with the mean shift [142] or the MEK algorithms [204, 242], and the application of morphological filters and blob detection [98, 233]. The most commonly used technique is template matching, where a template image, whether synthetically designed or extracted from a real image prior to tracking, is registered to an image, for which the markers have to be found. The marker positions are inferred from the minima of the distance measure function between the template and the image of interest [7, 31, 43, 122, 141, 140, 164, 169, 174, 199, 214, 234].
- **Measurement assessment/Correction**: Some measurement methods generate multiple candidates for the marker positions in the current image, and at this step the candidates are evaluated using some criteria to prune away false positives. Examples include such criteria as the combination of the template matching score[3] and relative marker positioning [122]; the template matching score and the size of the candidate markers [140]; the template matching score followed by the intensity and marker shape similarity between the candidate markers and the ones in the template, the smoothness of a marker trajectory, and the correspondence to a breathing pattern [214]; and the similarity between the measured and predicted marker shape/orientation, absolute position, and distance to the closest marker [142].

[2] In this case, the predicted marker position is the position of the marker from the preceding image. Sometimes, this has been referred to as the first-order predictor [142].

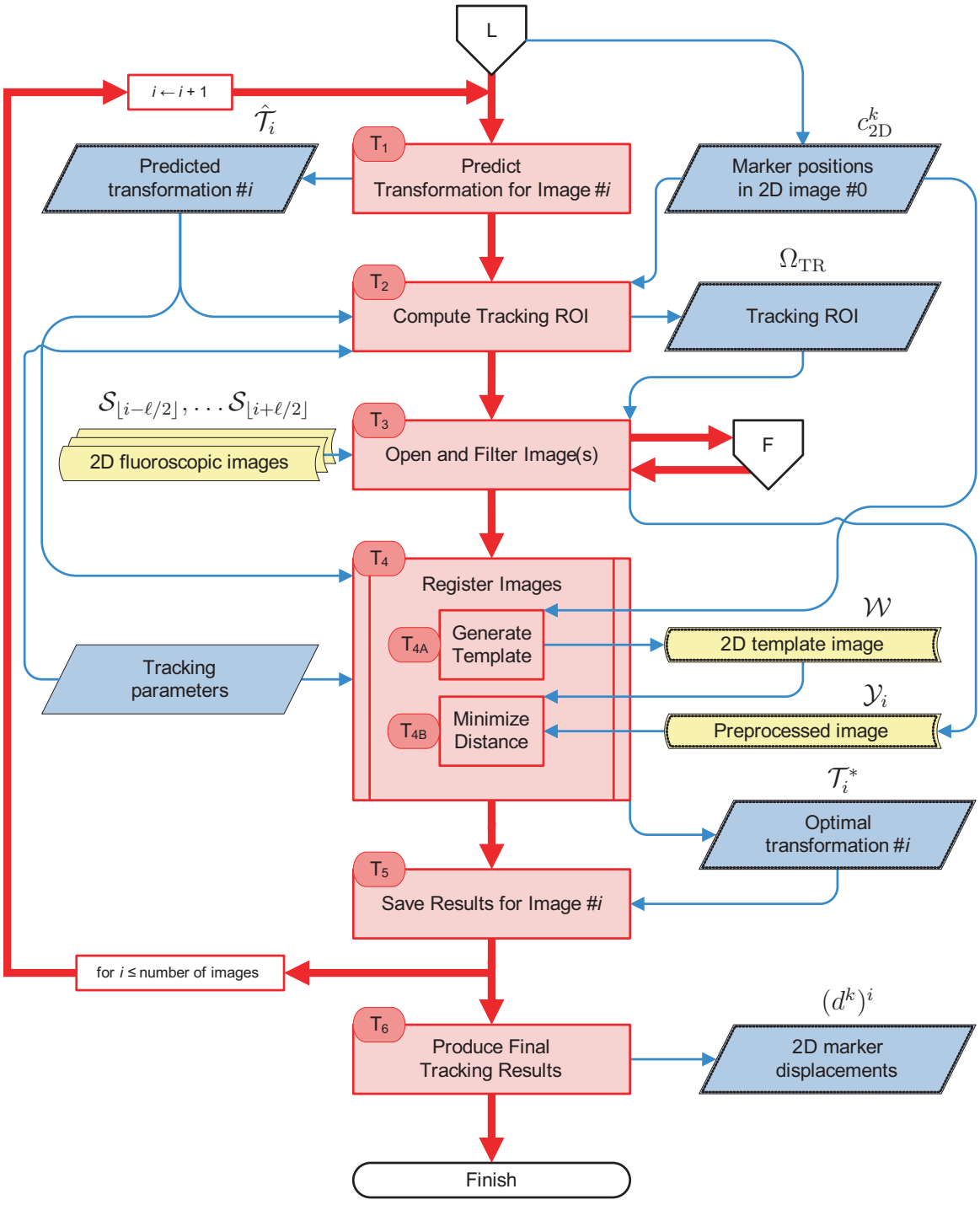
[3] The template matching score relates to the value of a distance measure function between a template and an image.

- **Tracking failure detection**: Some tracking systems rely on certain criteria to detect a tracking failure, in order to stop the treatment in real-time or report an error in post-processing. The reported methods mostly check for violation of the thresholds on the template matching score [196, 199, 214], marker speed, acceleration, smoothness of trajectory or its closeness to the predicted track [199, 214], and change in geometric relationship between the markers [214]. A detailed discussion was presented in Section 5.5.

In the remainder of this introduction, we present some of the assumptions and a scheme for the tracking module. Section 6.1 discusses the components of the 2D-2D image registration problem used in the measurement step of the tracking. The choice of the model for the prediction step is addressed in Section 6.2. After selecting an appropriate prediction model, we complete the design of the tracking module by providing the details about the optimization procedure used in registration in Section 6.3, and conclude with the discussion of the results in Section 6.4.

The design choices for our marker tracking procedure are based on a number of assumptions. We define organ *motion* observed in a fluoroscopic image sequence as the displacement of the markers in time relative to their positions in the tracking template. The *template* is created from one of the images, typically the first image of the sequence, for which the markers have been localized prior to tracking as described in Chapter 5. The displacements of the markers in a particular image relative to their positions in the template are measured by registering this image to the template. Hence, for each image in the sequence, a geometric transformation needs to be found that, when applied to the template, minimizes the chosen distance measure between the intensities in the template and the image.

Flowchart 6.1 shows a scheme for the marker tracking procedure. At the first step $\langle\!\langle \mathbf{T}_1 \rangle\!\rangle$, the prediction for the geometric transformation, $\hat{\mathcal{T}}_i$, is computed. Based on $\hat{\mathcal{T}}_i$ and the marker positions in the template, c^k_{2D}, a tracking ROI Ω_{TR}, over which the images will be registered, is defined at step $\langle\!\langle \mathbf{T}_2 \rangle\!\rangle$. Then, a fluoroscopic frame, or several frames in cases when temporal filtering is to be used, are selected from the sequence at step $\langle\!\langle \mathbf{T}_3 \rangle\!\rangle$. The control is passed to module $\langle\!\langle \mathbf{F} \rangle\!\rangle$ that applies the temporal and marker enhancement filters described in Chapter 4, and returns a preprocessed image $\mathcal{Y}_i$. At step $\langle\!\langle \mathbf{T}_4 \rangle\!\rangle$, the tracking template $\mathcal{W}$ is registered to $\mathcal{Y}_i$ over Ω_{TR} using the iterative optimization procedure described in Sections 6.1 and 6.3. The resulting transformation $\mathcal{T}^*_i$ is saved at step $\langle\!\langle \mathbf{T}_5 \rangle\!\rangle$, in order to be used later in computation of prediction for the subsequent images, and to produce the final tracking results. After the process is completed for all images in the sequence, the results are reported at the final step $\langle\!\langle \mathbf{T}_6 \rangle\!\rangle$ of the tracking procedure as the 2D marker displacements, $(d^k)^i$, relative to c^k_{2D}.



Flowchart 6.1: Block $\langle\langle \mathbf{T} \rangle\rangle$: Marker tracking in 2D fluoroscopic image sequence by 2D-2D image registration.

In this chapter, we complete the discussion of the design of the marker tracking system. In this last module, we combine and further develop digital filtering, template image generation, and registration and optimization techniques, discussed and proposed in Chapters 4 and 5. The use of the aforementioned novel approaches allows us to create a tracking system that is capable of accurate, reliable and computationally efficient marker tracking in noisy fluoroscopic image sequences of the prostate patients.

6.1 2D-2D Image Registration

The 2D-2D image registration problem in tracking is formulated in a similar way to the localization registration problem (5.3.10), where the parameters a^i of a geometric transformation have to be found that minimize some distance measure function $\mathcal{D}$ between the tracking template transformed by a^i, $\mathcal{W}_{a^i}$, and the (preprocessed) fluoroscopic image, $\mathcal{Y}_i$:

$$\begin{aligned}(a^*)^i &= \arg\min_{a^i} \mathcal{D}(\mathcal{W}_{a^i}, \mathcal{Y}_i),\\ \text{s.t. } &-a^{\max} \leq a^i \leq a^{\max},\end{aligned} \tag{6.1.1}$$

where $a^{\max}$ represents bounds on a^i and is discussed later in this section, and $\leq$ is an element-wise comparison. Based on the same considerations as in the 2D marker localization procedure (Section 5.3.3), the distance measure in tracking is minimized by applying the active set realization of the SQP optimization approach, beginning from some starting point a^0, that will be defined in Section 6.3.

Due to the assumption of the organ rigidity (Section 2.2), the particulars of the marker geometry (close to the isocentre) and limitations of our imaging system that was not able to detect the out-of-plane motion with sufficient accuracy in the marker localization module (Sections 5.3.2 and 7.3.3), we use the 2D rigid geometric transformation,[4] $\mathcal{T}^{2D}$, in our tracking procedure. The definition of $\mathcal{T}^{2D}$ can be found in Appendix B.2. The same transformation is applied to all markers, i.e., they are tracked together as a rigid structure.

The distance between the template and the reference images is measured by $\mathcal{D}_{\text{NNCC}}$, the negation of the normalized correlation coefficient (NCC, defined in Appendix C.4). This choice of the distance measure is based on a number of studies, where the NCC was successfully used for registration of the x-ray, fluoroscopic and CT/CBCT images and their projections [7, 26, 43, 59, 169, 174], and our own experience of using it in the proposed marker localization procedure to register the generated

[4] In the case of the 3D transformation, the template could be formed from the 3D CBCT image instead of the first 2D image of the sequence.

templates to the fluoroscopic images preprocessed by the marker enhancement filter (Section 5.2).

The vector $a^{\max}$ contains the bounds on the transformation parameters. The values we use for the bounds on the shifts, t_u and t_v, and rotation, γ, are the same as in the 2D marker localization procedure (Section 5.3.1), i.e., for prostate patients:

$$a^{\max} = (t_u, t_v, \gamma)^T = (20 \text{ mm}, 20 \text{ mm}, 10^\circ)^T. \tag{6.1.2}$$

They define a maximal possible displacement of the markers during a treatment session, whether it is between the acquisition of the CBCT image and the first image of the fluoroscopic sequence, or between any of the images within the sequence.

As mentioned in the introduction to this chapter, it is expected that the use of at least some form of prediction in a tracking application can be beneficial in terms of decreasing the processing time and improving the tracking reliability through reduction of the search space. In other words, given reasonable estimates of the marker positions in the incoming image, the search for the markers should only be performed in some close vicinity to these estimates. Hence, the motivation behind designing a separate procedure for tracking and not simply using the independent 2D marker localization in each image of the sequence is to create a method that is as reliable and accurate as the proposed localization technique, but is more computationally efficient. Although our tracking method is designed for post-processing, it should be remembered that it is aimed at analyzing large numbers of patient images. Therefore, spending less time and the ability to use the software on a conventional moderately equipped computer is always a value, and can open doors for real-time application of the methods proposed in this thesis.

The results of marker tracking in the preceding images of the sequence can be used to perform the prediction in the incoming images. The rate of our fluoroscopic image acquisition (5.5 Hz) is such that, for prostate patients, the scale of the observed marker motion between the adjacent frames is typically small. In particular, it is usually significantly smaller than the motion range defined by $a^{\max}$ parameters. This means that the tracking can be performed over a *tracking ROI* selected around the predicted marker positions, that is considerably smaller than the one defined by the parameters in $a^{\max}$ in (6.1.2). Therefore, we define another set of parameters, $b^{\max}$, that are used to compute the marker circular ROIs and the encompassing rectangular ROI, Ω_{TR}, according to the method proposed in Section 4.3, the same way as the bounds $a^{\max}$ are used to compute the localization ROIs, Υ_k and Ω_{LOC}. While the rectangular ROI in tracking is used for convenience of operating on images, such as selecting a subregion in the images, the distance measure computation is only performed over the circular ROIs. To reiterate, while $a^{\max}$ defines the maximal displacement of

the markers between the template and any of the images in the sequence, and represents bounds on the parameters of geometric transformation in the optimization problem (6.1.1), $b^{\max}$ is used to determine the ROIs used in registration.

The values in $b^{\max}$ depend on expectations regarding the maximal motion speed in a given data set. In the data set used for validation of the tracking procedure described in Section 7.4.1, the Quasar phantom was set to execute motion along the V direction. Therefore, based on the measurements of the maximal marker displacement and maximal change in shift and rotation parameter values between the adjacent images, we used $a^{\max} = (5 \text{ mm}, 25 \text{ mm}, 5^\circ)^T$, and $b^{\max} = (0.5 \text{ mm}, 8 \text{ mm}, 0.5^\circ)^T$. As defined in (6.1.2), $a^{\max}$ for the prostate patients is $(20 \text{ mm}, 20 \text{ mm}, 10^\circ)^T$. In our patient testing data set, described in Section 7.4.1, that includes 25 fluoroscopic sequences of three prostate patients, the maximal changes in the parameter values between the adjacent frames were 1.5 mm, 0.5 mm and 2.7° for the U and V shifts, and the rotation angle, respectively. The maximal 2D displacement between the markers was 2.44 mm. As we expect that the prediction is at least as good as using the marker positions from a preceding frame (zeroth-order prediction), the values given above can be used to determine $b^{\max}$. Due to the expectation that other patients can potentially have more motion, we decided to use $b^{\max} = (5 \text{ mm}, 5 \text{ mm}, 5^\circ)^T$.

The ROI shrinkage introduced by using $b^{\max}$ instead of $a^{\max}$ can reduce the chances of finding in the ROI the measurements of anatomical features that possess significant contrast and are of a width similar to the markers. Therefore, in contrast to the localization procedure, in tracking it may be possible to simplify or omit altogether the preprocessing of the images, such as the application of the marker enhancement filter (MEF) to fluoroscopic images (Section 4.4.1), and hence reduce the computational time. We compare two approaches:

- **Registration using interpolation, without preprocessing**: In this approach, the MEF-preprocessing, as well as the specific template image generation approaches (Section 5.1), are not used. The registration is performed directly between the image and the template, where the template is the first image of the sequence. In practice, the images are discrete, and the intensity function is not defined between the grid points. The application of the geometric transformation to the template requires to infer the intensity values between the grid points in order to compute the distance between the transformed template and the reference images. These intensity values can be found by using some form of *interpolation*, where some of the most commonly used interpolation schemes are the nearest neighbour, linear, cubic, and cubic spline [150, 69].
- **Registration using marker enhancement preprocessing**: Another realization of 2D-2D image registration is similar to our 2D-3D registration method

described in Chapter 5. In the process of registration, the template is generated using the marker splat (MS) method (Section 5.1.3), modified for the 2D as follows. Let $\tilde{c}^k_{2\mathrm{D}}$ be the position of the marker k that is obtained by applying a transformation $\mathcal{T}^{2\mathrm{D}}$ to the marker position in the template image found by 2D marker localization, $\tilde{c}^k_{2\mathrm{D}} = \mathcal{T}^{2\mathrm{D}}(a, c^k_{2\mathrm{D}})$. Then, the modification of the MS DRR (Equation (5.1.9)) for the 2D (2D MS) can be computed as:

$$\mathcal{W}_k(q) = \mathcal{G}(q - \tilde{c}^k_{2\mathrm{D}}, \sigma), \tag{6.1.3}$$

where the $\mathcal{W}_k$ for individual markers are then summed up, and rescaled to take the range $[0, 1]$, resulting in $\mathcal{W}_a$ (6.1.1). The reference image is preprocessed prior to registration by the MEF or MOF, in order to amplify the markers and suppress the background, and to obtain an image that has similar intensity and structural characteristics to the generated template. Hence, interpolation is not required, and the template is quickly recomputed for every new set of transformation parameters a.

To compare these two approaches, we tested our Matlab implementation on several fluoroscopic image sequences of the Quasar motion phantom. The detailed descriptions of the data and experiment are presented in Section 7.4.2. The results suggest, that though the registration using interpolation does not require an application of the MEF, it may lead to much higher computational cost compared to the independent 2D localization in every image of the sequence. The costs are similar for the nearest neighbour, up to 9 times higher for the linear, 16 times for the cubic, and 7 times for the cubic spline interpolation schemes. The application of the MEF to the reference image and generation of the template by the 2D MS method offered approximately 1/5 reduction in time compared to the independent 2D localization. Further, we explored the fact that the MEF combines two images, produced by the magnitude-only and magnitude-and-ratio methods, where the first one aims to amplify and preserve the marker shapes while the second helps to suppress non-marker features that are likely to be found in a relatively large localization ROI (Section 4.4.1). Due to the ability to predict the 2D marker positions and, consequently, to use the tracking ROI that is smaller than the localization ROI, we were able to successfully substitute the application of the full MEF by its reduced version, the MOF method, which is based on the magnitude-only image, and is described in Section 4.4.5. This brought a significant reduction in computational time by being almost 6 times faster than the independent 2D localization method.

We suggest that the reasons for significant computational savings of the approach using the MOF preprocessing comparing to the interpolation-based techniques are the following:

- The MS template image generation method is characterized with very short computational time due to its simplicity, and a limited number of points involved in the computation: The markers are generated as the Gaussians centered in the transformed marker positions. Interpolation, on the other hand, needs to be performed over all points of the image inside the registration ROI, resulting in higher costs associated with data access and computation.
- The MOF preprocessing requires insignificant additional time[5] and creates an image similar in characteristics to the template, making registration reliable and fast.

In addition to efficiency comparison, we have also compared the tracking accuracy of the approaches to the independent 2D marker localization. The median fiducial registration error (FRE, Equation (7.3.19)) was 0.0002 and 0.0003 mm for the MOF and MEF methods, respectively, 0.02 mm for the nearest neighbour, and 0.15–0.16 mm for the linear, cubic, and cubic spline interpolation. Although all methods showed very good (subpixel) tracking accuracy for almost all images, the application of the MEF or MOF preprocessing coupled with the use of the 2D MS template generation method provided a significant improvement in the computational efficiency (Section 7.4.2). Therefore, we use this approach in the 2D-2D image registration procedure of our tracking module. The next section discusses the selection of an appropriate prediction model.

6.2 Prediction Models

As mentioned in the introduction to this chapter, while some methods perform tracking using the independent 2D localization of the markers in each image of the sequence, others take advantage of some form of prediction to assist and guide the tracking process. The main purpose of prediction in our post-processing application is the reduction of the tracking ROIs in order to save computational time and increase the reliability of tracking.

Although it is possible to observe relatively regular small prostate motion caused by respiration, it is well accepted that most prostate moves can be characterized as incident and irregular. Therefore, though we anticipate that a prediction model can serve well in terms of providing the possibility for keeping the tracking ROIs small, it may be rather challenging to predict the marker positions accurately.

The majority of the prediction models reported in the literature are some form of linear prediction [195, 40, 234, 214, 123, 79, 140, 142]. In general, linear models use

[5] In the experiment with the Quasar motion phantom described in Section 7.4.2, the MOF preprocessing took approximately half a second per sequence of 164 frames.

marker or target positions found in the few preceding images, and a linear model, to make a prediction in the current image. The following linear prediction methods are described in the literature: zeroth-order (ZO) prediction [140, 142], linear extrapolation (LE) [195], linear regression (LR) [195, 183, 214, 234], and linear adaptive (LA) models [221, 40, 79]. Also used are artificial neural networks (ANN) [195, 123], support vector machines (SVM) [123], Kalman filter (KF) [195, 186, 130, 216], particle filter [186, 200], and polynomial models [123].

Sharp *et al.* [195] compared different prediction models for motion tracking in 14 lung tumour cases with peak-to-peak motion larger than 8 mm. The prediction accuracy was evaluated by the magnitude of the root mean squared (RMS) error between the known and predicted tumour positions, averaged for all patients. For imaging frequencies of 1–30 Hz and latencies of 33, 200, and 1,000 ms, the LR model gave similar or better results in comparison to ANN, and consistently outperformed the KF. In particular, for imaging and prediction conditions similar to ours, i.e., imaging rate of 5.5 Hz and prediction for the next incoming image, which approximately corresponds to a latency of 200 ms in [195], the RMS error for no prediction was approximately 3.3 mm, improved by the KF and ANN to 3 mm, and by LR to 2.5 mm.[6]

Lin *et al.* [123] compared the LR, 2^{nd}-degree polynomial, ANN, and SVM models for tracking certain features in 10 fluoroscopic image sequences of 10 lung cancer patients with the range of motion from 4 to 25.5 mm.[7] All methods provided comparable results with the mean prediction errors of approximately 1.1 mm for the LR, polynomial, and SVM, and 1.05 mm for the ANN. After analyzing errors in the training and testing data sets, the authors suggested that the 2^{nd}-degree polynomial may be over-fitting the data. Also, the ANN method may be more reliable, and its maximal mean prediction error among different patients is smaller (2 mm for the ANN comparing to 3 mm, 3.05 mm, and 3.45 mm for the LR, polynomial, and SVM, respectively).

Krauss *et al.* [109] compared the performance of respiratory motion prediction based on the LR, ANN, kernel density estimation (KDE), and support vector regression (SVR) for various sampling rates and system latencies ranging from 0.2 to 0.6 seconds. The prediction models were evaluated on 12 3D lung tumour motion traces. When averaging over all sampling rates and latencies, prediction errors normalized to errors of using no prediction for the ANN, SVR, LR, and KDE were 0.44, 0.46, 0.49, and 0.55, respectively, i.e., a two-fold improvement in positioning accuracy was observed.

Rottmann *et al.* [183] used the LR to perform prediction for dynamic multi-

[6] Data is inferred from Figure 8 in [195].

[7] The results in [123] are reported in pixels, and are translated to millimeters here for consistency of discussion.

leaf collimator system with the latency of around 230 ms and imaging frequency of 12.86 Hz. Image sequences reproduced the motion of a phantom that was programmed to use the real motion target trajectories of three lung patients. The predictor was trained on the first 100 target positions of each image sequence. The RMS error between estimated tumour position and beam aperture position was between 0.6 ± 0.6 mm and 0.9 ± 0.9 mm for different traces.

Verma *et al.* [223] published a survey of real-time tumour motion prediction methods used in image-guided radiation therapy. Comparison of the methods reported by different authors reveals trends similar to the results discussed above: For the 200 ms latency, (i) the improvement in accuracy comparing to no prediction ranges from no improvement to around 2.5 times, and (ii) the errors in target position prediction are similar for different variations of the ANN, KF and linear models.

Based on the reviewed literature, we concluded that the use of more sophisticated models in our tracking application, such as the KF, ANN, SVM, or KDE, is not likely to offer significant improvement in prediction compared to one of the simpler linear models, for example, linear extrapolation or linear regression, discussed below. To select the model and its parameters, we conducted an experiment using fluoroscopic sequences of the Quasar motion phantom (the details are provided later in this section). In the following, the vector of predicted parameters in image i is denoted by $\hat{a}^i$, and the optimized parameters by $(a^*)^i$. We compared the following models:

- **No prediction**: No prediction means that the predicted positions of the markers in every image of the sequence are the same as the positions localized in the first image. As the tracking parameters in image i refer to the displacement between the markers in this image and in the template (created from the first image), the predicted parameters are:

$$\hat{a}^i = (0, 0, 0)^T. \tag{6.2.4}$$

  No prediction is used as a baseline for comparison and allows to measure a motion range of the markers during acquisition of the whole fluoroscopic sequence.
- **Zeroth-order prediction (ZO)** [142, 140]: In this method, the optimized parameters of the geometric transformation resulting from the registration between the preceding image and the template are used as the predicted parameters in the current image:

$$\hat{a}^i = (a^*)^{i-1}. \tag{6.2.5}$$

  While it can be used on its own [140], ZO prediction can also be combined with gantry rotation information or other marker motion [142].
- **Linear extrapolation (LE)** [199, 195]: This method assumes a constant change in the parameters and uses the optimized parameters from the two pre-

ceding images of the sequence:

$$\hat{a}^i = 2(a^*)^{i-1} - (a^*)^{i-2}. \tag{6.2.6}$$

- **Linear regression (LR)** [195, 214, 123, 183, 221]: The predicted parameters are computed as a linear combination of the optimized parameters from n preceding images, $(a^*)^{i-1}, \ldots, (a^*)^{i-n}$:

$$\hat{a}^i = \beta_0 + \beta_1(a^*)^{i-1} + \ldots + \beta_n(a^*)^{i-n}, \tag{6.2.7}$$

  where the coefficient vector β is obtained beforehand from training. The number of preceding images, n, used in the LR model is typically defined from experiments and depends on the characteristics of the motion (speed, regularity) and image acquisition (frame rate). Our approach to computing the coefficients β is explained in Section 7.4.3.

  The use of LR model [195, 183] can be extended by combining the LR prediction with the breathing pattern [214]. It can also be used to find a correlation between the marker motion observed in images produced by the kV and MV imagers [234].
- **Linear adaptive (LA)** [40, 221, 79]: It can be argued that the nature of motion can change over time, and as such, the coefficients obtained from the training data can become obsolete and have to be corrected. For every incoming image $\mathcal{Y}_i$, the update can be done based on the closest in time available data, i.e., by minimizing the squared error for the preceding image $\mathcal{Y}_{i-1}$ between the predicted parameters, $\hat{a}^{i-1}$, and the optimized parameters, $(a^*)^{i-1}$, obtained from the registration between $\mathcal{Y}_{i-1}$ and the template $\mathcal{W}_{a^{i-1}}$.

While the geometric parameters $\hat{a}$ for no prediction, ZO, and LE models can be computed in a straightforward way, the LR and LA models typically have to be trained beforehand on testing image sequences in order to obtain the coefficients $\beta = (\beta_0, \ldots, \beta_n)^T$ to be used in Equation (6.2.7). A detailed discussion on how to compute β is provided in Section 7.4.3. In the LA model, the coefficients are updated for every new incoming image. The update strategies are described in [40, 221, 79], and are based on the idea of minimizing the prediction error in the preceding images, for which the marker positions have been already found. In this case, the elements of β have a potential of significant change from image to image, especially in the scenarios when the markers have been nearly stationary and then make a large rapid move. Therefore, in the proposed application, the prediction can become infeasible, i.e., the suggested geometric parameters $\hat{a}$ can fall outside of their feasible region, limited to $a^{\max}$, which we indeed observed in our experiments. Hence, this approach

requires careful application, tuning of parameters or some correction strategy, and even with the correction, the prediction can cause misregistration between the image and the template as the starting point for the optimization procedure can still be too far from the solution. As such, the LA model is not a good choice for our tracking procedure.

At the same time, it may be reasonable to apply β obtained from a training data set to all new images, such as assumed by the LR model. As noted by Krauss *et al.* [109], the use of the same parameters β computed from model parameter optimization on a patient-average basis resulted in successful application of the LR model to the patient-specific motion patterns.

Similarly to Krauss *et al.* [109], we compute β from various marker traces of the Quasar motion phantom in the experiment described in Section 7.4.3. The data set consists of 23 fluoroscopic sequences, where the phantom is performing several types of motion, such as sinusoidal in time (linear in direction), simulated patient breathing traces, and linear motion (for detailed description, see Section 7.4.1). We believe that this set of trajectories includes a fair representation of possible ranges and speeds of marker motion of the prostate patients, plus a number of trajectories of even larger speeds and amplitudes, more typical of the lung or liver patients.

The same data set was used to compare the ZO, LE, and LR prediction models in terms of fiducial prediction error (FPE), defined in Section 7.4.4. The temporal filtering has not been used for phantom images due to sufficient marker contrast.

As can be seen from the results presented in Section 7.4.4, the ZO and LE models are characterized with very similar FPE, and improve most of the statistical indicators 10-fold and maximal FPE around 3.5 times in comparison to no prediction. The FPE results for the LR model with different number of images are very similar, which is explained by the fact that the corresponding coefficients β_i are similar regardless of the number of preceding images used, and the two largest coefficients are assigned to the geometric parameters of the two preceding images (see Section 7.4.3). The median FPE of the LR approach is similar to that of the ZO and LE models (0.04 mm difference that corresponds to approximately 1/6 of a pixel size), however, the maximal FPE is improved by around 1–1.5 mm (approximately 4–6 pixels).

The results reported in Section 7.4.4 showed that all considered models reduced the FPE significantly comparing to no prediction. The LR models were better in reducing the maximal FPE than ZO and LE. Therefore, we select the LR model based on three preceding images (LR-3) as a predictor in the proposed marker tracking module, as it provides slightly better results than the rest of the LR models. The coefficient vector used was $\beta = (0, 1.99, -1.23, 0.23)^T$. We have also validated the models on the 25 fluoroscopic image sequences of the patient testing data set in

Section 7.4.5, and our choice of the LR-3 was confirmed to be reasonable.

6.3 Optimization

The analysis of the results presented in Sections 6.2, 7.4.4, and 7.4.5 shows that for the LR-3 model, the maximal FPE for the fluoroscopic image sequences of the Quasar motion phantom reached as large as 5.39 mm (approximately 21 pixels). For the patient data sets the typical maximal FPE was below 0.36 mm (slightly larger than 1 pixel) and several outliers reached 1.63 mm (more than 6 pixels). With such FPE values, and taking into consideration the fact that other patients may have a larger range of motion, the use of the predicted geometric parameters $\hat{a}$ as the optimization starting point a^0 in the 2D-2D image registration procedure (6.1.1) may cause the optimization method to become trapped in a local minimum or to fail finding the direction of the decrease of the distance measure function, which can result in misregistration or registration failure. Indeed, our experiment showed that for the 23 Quasar image sequences described in Section 7.4.1 (3,723 images in total), the misregistration or failure happens in 12.35% of images.

Fortunately, the problem of misregistration can be easily overcome by correcting the starting point. We base our strategy on the results of the experiments performed to make design choices for the 2D marker localization procedure (see Section 5.3.3). In particular, there we compared different methods for computation of the optimization starting point for the 2D-3D image registration, and found that the normalized cross-correlation (NCC) on reduced resolution with multiple templates (different rotation angles) was the most reliable and computationally efficient method. As 2D-2D registration in the tracking procedure operates on images similar to the ones undergoing 2D-3D registration in the 2D marker localization, we expect that the NCC approach can be as effective. As described in detail in Appendix D.5.1, the starting point a^0 is computed as the combination between the prediction $\hat{a}$ and the point where the NCC between the template transformed by $\hat{a}$ and the image reaches its maximum. However, as the tracking ROI is smaller than the localization ROI, and not much rotation is expected between the consecutive images of the sequence, the procedure can be further simplified by using the shift-only NCC on the original resolution (single template, no rotation). For our Matlab implementation executed on the system described in Appendix A, the use of the NCC on average took one additional second per whole image sequence in comparison to no NCC ($a^0 = \hat{a}$) when tested on the Quasar motion phantom image sequences described in Section 7.4.1, but increased the registration success rate to the full 100%.

6.4 Discussion

This chapter described the design of the marker tracking procedure, and concluded the whole marker localization and tracking approach proposed in this thesis. In Section 6.1, we defined a 2D-2D image registration problem used in tracking, and discussed some of its components, such as the distance measure, the optimization method, the geometric transformation, and the bounds on the geometric transformation parameters, all of which were selected based on our findings about 2D-3D image registration problem used in 2D marker localization (Section 5.3). We discussed the possibility of reduction of the tracking ROI in comparison to the localization ROI due to a smaller displacement expected between the markers in the consecutive images of a sequence compared to the expected displacement between the CBCT and the fluoroscopic sequence. This had two important implications. First, there was a necessity to select another set of parameters that defined this new ROI. The second implication is explained below.

We have compared two 2D-2D image registration approaches: (i) the direct registration of the images to the template (first image of the sequence) by using a traditional registration technique involving interpolation and (ii) the preprocessing of the images in the sequence using the MEF with the subsequent registration to the template formed from the 2D marker positions localized in the first image of the sequence. As a basis for comparison, we also performed an independent 2D marker localization in each image of the sequence. The details for this experiment are provided in Sections 6.1 and 7.4.2. While the marker positioning accuracy was good for all methods (higher for the MEF), the use of the nearest neighbour interpolation incurred a lower computational cost than the independent 2D marker localization. However, other methods using linear, cubic, and cubic spline interpolation schemes incurred similar cost for some sequences and substantially higher costs for most. The second approach combined the use of the MEF with the marker splat template generation method modified for the 2D (2D MS, summarized in Algorithm D.17). This approach led to some reduction in running time compared to the independent 2D marker localization. Recall that the MEF combines two images: magnitude-only and magnitude-and-ratio, where the first one aims to amplify and preserve the marker shapes while the second helps to suppress non-marker features that are likely to be found in a relatively large localization ROI (Section 4.4.1). Since the tracking ROI is smaller, and is expected to include a smaller number of anatomical features that may be amplified by the filter, we exploited the idea of using a variation of the MEF, the magnitude-only filter (MOF, defined in Section 4.4.5), in the tracking. This is a second implication of selecting a smaller ROI for tracking compared to localization ROI mentioned in the previous paragraph. In addition to providing a better accuracy

than the interpolation-based methods, the use of the MOF combined with the 2D MS template generation method led to a significant reduction in the computational cost, and thus, we decided to use it in our tracking procedure.

Further, we reviewed motion prediction models presented in the literature in Section 6.2, and selected a number of promising approaches, such as zeroth-order predictor (ZO), linear extrapolation (LE), and linear regression (LR). The application of the models to testing data set of the Quasar motion phantom described in Section 7.4.1 revealed that the LR based on three preceding images is the most suitable model for our typical data, though ZO and LE models also performed well. The experiments and the results are described in Sections 7.4.3 – 7.4.5.

Then, we discussed the choice of the starting point for the optimization procedure in Section 6.3. Due to similar components of the optimization problems for 2D-2D and 2D-3D image registration, tracking failure detection and prevention procedures are the same (see Section 5.5).

Finally, the tracking accuracy results are presented in Section 7.4.6. For both patient and Quasar data sets described in Section 7.4.1, the proposed system demonstrated subpixel and submillimeter tracking accuracy, with extremely small tracking errors for most image sequences. As such, it can be implemented for clinical use. For the Quasar data set, no correlation was found between the tracking error and the trajectory, speed, and magnitude of the motion, though the maximal errors tended to be slightly higher for the sequences with higher phantom motion speed.

Experiments for evaluating the accuracy and performance of the tracking module are collected in Section 7.4, and implementation details and algorithms are listed in Appendix D.5. Chapter 8 provides an example of the results that can be obtained for patient images using the proposed marker localization and tracking methods described in Chapters 3 – 6 and implemented in the Gryphon software module.

Part II

Validation and Case Studies

Chapter 7

Selection of Parameters and Validation of Methods

This chapter provides descriptions of experiments for selecting some data-dependent parameters for the proposed marker localization and tracking procedures, and includes a detailed description of the testing data sets. The greatest part of the chapter is occupied by the validation results, such as definitions of errors and error estimation procedures, assessment of accuracy of the methods, and their success rates in application to testing data.

As a general rule, we report the quantitative results, such as error distributions, in terms of their minimal, median, and maximal values, as well as their 25^{th} and 75^{th} percentile, both graphically and numerically. The corresponding notation for tables and figures is explained in Table 7.1. This is done to achieve the reporting that is compact enough to be presented in this thesis, but does not place too many assumptions on the distributions. However, a widely accepted practice, in particular, in radiation therapy community, is to assume the normal distribution of the errors. Therefore, to provide an easy comparison to the literature, where applicable, we also report the results in terms of their mean and standard deviation values.

7.1 Marker Segmentation and Modelling in 3D Space

7.1.1 Testing Data Sets for Marker Segmentation and Modelling

Table 7.2 summarizes details on all data sets used for validation of 3D marker segmentation and modelling procedure. Radio-surgery verification phantom (RSVP, de-

Table 7.1: Notation for distributions of values in figures and tables.

Statistical value	Table notation	Figure symbols
median	Med	*dotted circle*
minimum, maximum	Min, Max	*thin bar*
25^{th}, 75^{th} percentile	25%, 75%	*thick bar*
outliers	Outliers	*empty circles*
mean, average	Mean	
one standard deviation	Std	

scribed in Section 2.3.1) was placed in nine different positions, and CBCT projections of 512×512 pixels were acquired at each position. Then, the clinical software (XVI) used the projections to reconstruct 3D CBCT images with 0.5, 1, 2, and 4 mm voxel sizes. Clinically, images with 0.5 and 1 mm voxels are most commonly used but for some sites the use of 0.25 mm voxels can be desirable. As it was not possible to reconstruct an image with 0.25 mm voxels from the originally acquired CBCT projections, and RSVP was not available at the time, we acquired another set of 5 CBCT images (with different shifts along the Y axis). This was done with the same phantom insert containing seven gold cylindrical fiducial markers 3×0.8 mm but placed inside the Quasar phantom described in Section 2.3.2. As a result, the CBCT reconstructions with 0.25 mm voxels were obtained from the CBCT projections of 1024×1024 pixels.

In addition to all the images described above that were reconstructed using the `float` data type, we also had one data set reconstructed from the same set of 512×512 pixel projections of the RSVP with the `short` setting. Among those, one image had 1 mm voxels, and the remaining eight had 0.5 mm voxels.

Justifications of parameter choices were also done for the patient testing data set, which included measurements of three patients. We had 74 CBCTs for patients 1 and 3, and 75 CBCTs for patient 2, all of which had 1 mm voxels. For patient 1, we also had 6 CBCTs with 2 mm voxels. All patients had three gold cylindrical fiducial markers 3×0.8 mm.

7.1.2 Threshold Flexibility

This section presents the results of thresholding the CBCT images with different values of the flexibility parameter, κ, defined in Section 3.1.1. The value of κ influences the number of the segmented voxels, n_{vx}. The model-based thresholding proposed in Section 3.1.1 was applied to all data sets described in Table 7.2. Table 7.3 lists the thresholding success rates, defined as the percentage of successfully thresholded

Table 7.2: Testing data sets for validation of 3D marker segmentation and modelling.

Data set	Measure-ments of	Voxel size, mm	Data type	Number of images	Total number of markers
F 0.25 mm	Quasar	0.25	`float`	5	35
F 0.5 mm	RSVP	0.5	`float`	9	63
F 1 mm	RSVP	1	`float`	9	63
F 2 mm	RSVP	2	`float`	9	63
F 4 mm	RSVP	4	`float`	9	63
S 0.5 mm	RSVP	0.5	`short`	8	56
S 1 mm	RSVP	1	`short`	1	7
P1 1 mm	Patient 1	1	`short`	74	222
P2 1 mm	Patient 2	1	`short`	75	225
P3 1 mm	Patient 3	1	`short`	74	222
P1 2 mm	Patient 1	2	`short`	6	18

Table 7.3: Thresholding success rates depending on the choice of the flexibility parameter κ.

Data set	Thresholding success rates, % for threshold flexibility parameters:				
	$\kappa = 0$	$\kappa = 0.25$	$\kappa = 0.5$	$\kappa = 0.75$	$\kappa = 1$
F 0.25 mm	100	100	100	100	100
F 0.5 mm	92	100	100	100	100
F 1 mm	70	100	100	100	100
F 2 mm	17	43	75	97	100
F 4 mm	8	8	30	97	100
S 0.5 mm	100	100	100	100	100
S 1 mm	71	100	100	100	100
P1 1 mm	67	100	100	100	100
P2 1 mm	61	100	100	100	100
P3 1 mm	58	88	100	100	100
P1 2 mm	11	67	100	100	100

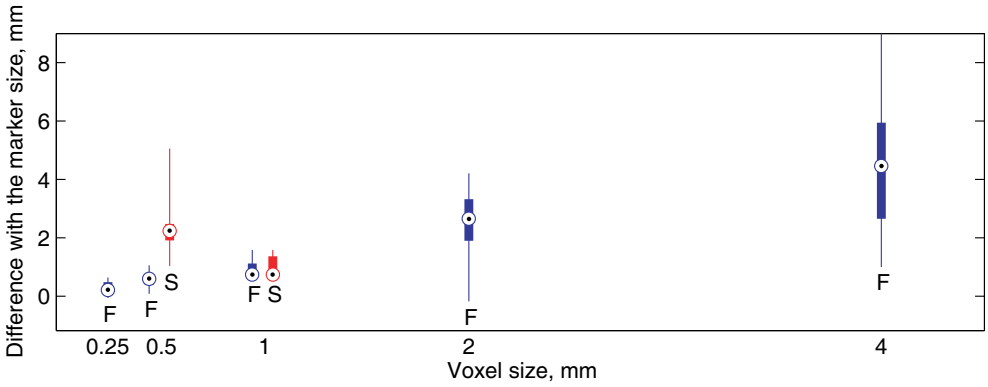


Figure 7.1: The difference between cluster diameters, $\mathrm{diam}(\hat{\mathrm{M}}_k)$, and the marker size, δ_{m}, for Quasar phantom (0.25 mm voxels) and RSVP (0.5 to 4 mm voxels) data sets described in Table 7.2. Table 7.4 provides corresponding numerical values.

markers. Success was registered if the marker consisted of at least two voxels, as this is the minimal number of voxels that allow to estimate the size of the segmented feature, which is required at step $\langle\langle \mathbf{S}_3 \rangle\rangle$ of the segmentation procedure.

The results in Table 7.3 suggest that among values $\kappa = \{0, 0.25, 0.5, 0.75, 1\}$, the successful thresholding for all data sets is achievable with $\kappa = 1$. As κ decreases, the success of thresholding drops. Increasing of κ beyond 1 could not provide any improvement but caused more non-marker points to be included into set Θ. Therefore, in this thesis, we use $\kappa = 1$. If it is known that no images with the voxel sizes larger than 1 mm are included into data sets, the use of $0.5 \leq \kappa < 1$ is reasonable, and can decrease the number of thresholded non-marker points, thus reducing the burden on the correction procedures in steps $\langle\langle \mathbf{S}_3 \rangle\rangle$ and $\langle\langle \mathbf{S}_4 \rangle\rangle$ of the proposed 3D marker segmentation procedure.

7.1.3 Marker Set Size Margins

Marker set size margins, ϵ_ℓ and ϵ_{u}, introduced in Section 3.1.3, Equation (3.1.12), define a range of acceptable set diameters around the known marker size, δ_{m}. The clusters whose diameters are within this range are accepted as the candidate marker sets. In order to determine appropriate values for ϵ_ℓ and ϵ_{u}, we computed the diameters of the clusters that corresponded to the markers in all data sets described in Table 7.2. Due to the fact that in the estimation of the number of voxels, n_{vx}, that should constitute markers (Equation (3.1.6) in Section 3.1.1) the height of the marker is extended by $\epsilon_2 = 2\,\kappa\,\delta_{\mathrm{vx}}$, where δ_{vx} is a voxel size, and $\kappa = 1$ (Section 7.1.2), it is expected that the difference between the diameters of the candidate marker sets,

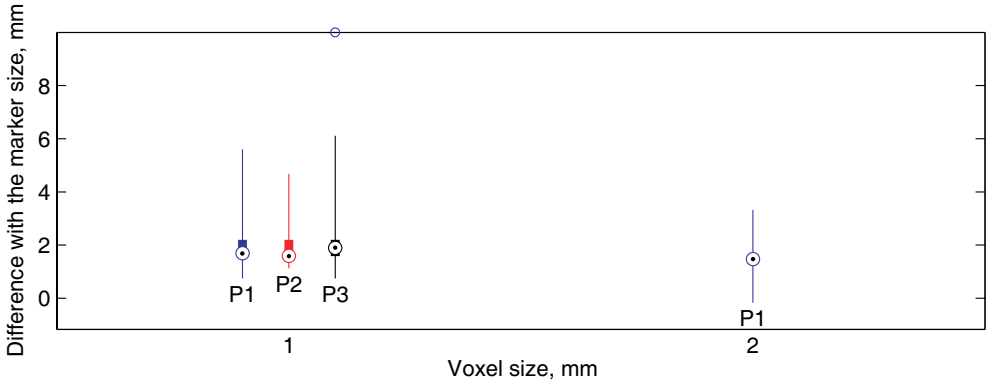


Figure 7.2: The difference between cluster diameters, $\mathrm{diam}(\hat{\mathrm{M}}_k)$, and the marker size, δ_{m}, for the CBCTs of three patients with 1 and 2 mm voxels, where the data sets are described in Table 7.2. Table 7.4 provides corresponding numerical values.

Table 7.4: The difference between cluster diameters, $\mathrm{diam}(\hat{\mathrm{M}}_k)$, and the fiducial marker size, δ_{m}, for phantom and patient CBCT images. Figures 7.1 and 7.2 provide corresponding graphical interpretation.

Data set	$\mathrm{diam}(\hat{\mathrm{M}}_k) - \delta_{\mathrm{m}}$, mm					
	Min	25%	Med	75%	Max	Outliers
F 0.25 mm	−0.06	0.13	0.22	0.48	0.64	–
F 0.5 mm	0.08	0.39	0.61	0.74	1.06	–
F 1 mm	0.61	0.74	0.74	1.12	1.58	–
F 2 mm	−0.17	1.9	2.66	3.32	4.21	–
F 4 mm	1	2.66	4.46	5.94	9	–
S 0.5 mm	1.03	1.91	2.24	2.48	5.06	–
S 1 mm	0.61	0.74	0.74	1.37	1.58	–
P1 1 mm	0.74	1.58	1.69	2.2	5.6	–
P2 1 mm	1.12	1.58	1.58	2.2	4.68	–
P3 1 mm	0.74	1.58	1.9	2.2	6.11	10
P1 2 mm	−0.17	1.47	1.47	1.47	3.32	–

$\text{diam}(\hat{\text{M}}_k)$, and δ_{m} should be around $2\,\delta_{\text{vx}}$.

The results for the Quasar and RSVP presented in Table 7.4 and Figure 7.1, with the exception of the "S 0.5 mm" data set, support the hypothesis given in the previous paragraph: The median and maximal differences between the cluster diameters and the marker size are about δ_{vx} and $2\,\delta_{\text{vx}}$, respectively. For the patients, results for which are presented in Table 7.4 and Figure 7.2, the maximal differences for the "P1 2 mm" data set are within $2\,\delta_{\text{vx}}$ but all 1 mm voxel data sets demonstrate larger maximal values, similar to the "S 0.5 mm" data set. The reason for this is the presence of the streak artifacts associated with the `short` reconstruction data type discussed in detail in Section 3.1.3. Although most of the values are smaller, we assign $\epsilon_{\text{u}} = 11$ mm, in order to accept all clusters corresponding to true markers, including the outliers, such as 10 mm for the "P3 1 mm" data set. Typically, the non-marker clusters that are accepted due to these loose size margin are positioned further away from the isocentre, and thus are removed by correction at step $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$ of the segmentation procedure. As for the lower margin, none of the data sets had the difference between $\text{diam}(\hat{\text{M}}_k)$ and δ_{m} being below -0.17 mm. Therefore, we assign $\epsilon_\ell = 0.5$ mm.

7.1.4 Correction Procedures and Segmentation Success Rates

The application of the marker set segmentation technique described by steps $\langle\!\langle \mathbf{S}_1 \rangle\!\rangle$ – $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$ presented in Sections 3.1 and 3.3 provided a 100% marker detection success in almost all data sets described in Table 7.2, i.e., 100% of true positives (TP) and 0% of false positives (FP). The only exception was the "F 4 mm" data set, which yielded 83% of TP and 2% of FP. For the nine images in that set, and seven fiducial markers visible in the images, the TP/FP numbers of the detected markers were 6/1, 5/0, 7/0, 5/0, 5/0, 7/0, 5/0, 7/0, 5/0. In the first image, one non-marker set was close to the isocentre and of an acceptable size. In all images, in which TP = 5, two neighbouring markers were rejected because they occupied the neighbouring voxels, and hence were recognized as one large cluster. This means that if the distances between the neighbouring fiducial markers in the phantom were larger than 10 mm, the detection success rates for 4 mm voxel images could be higher as well. For example, average inter-marker distances for the patient data sets described in Table 7.2 ranged from 20 to 36 mm.

Among all non-marker clusters, some were removed by violating the size requirement (Equation (3.1.12) in Section 3.1.3, step $\langle\!\langle \mathbf{S}_3 \rangle\!\rangle$), and others due to their location further away from the isocentre in cases when the number of candidate marker sets, $\hat{n}$, was larger than the known number of the fiducial markers, n (Section 3.3, step $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$). The number of clusters removed by either of the methods varies for different

Table 7.5: The average number of the non-marker clusters removed per image at steps $\langle\langle \mathbf{S}_3 \rangle\rangle$ and $\langle\langle \mathbf{S}_4 \rangle\rangle$ of the marker segmentation procedure.

Data set	Number of clusters removed, per image	
	Step $\langle\langle \mathbf{S}_3 \rangle\rangle$	Step $\langle\langle \mathbf{S}_4 \rangle\rangle$
F 0.25 mm	0	0
F 0.5 mm	0	0
F 1 mm	0	0.78
F 2 mm	0	0.78
F 4 mm	1.89	0.33
S 0.5 mm	0	0
S 1 mm	0	1
P1 1 mm	0	0.01
P2 1 mm	0.76	0.99
P3 1 mm	0.42	0.19
P1 2 mm	1.33	2.67

data sets, however, as it can be seen in Table 7.5, both correction procedures are important for reliable segmentation. As the number of images in the data sets was different, in the table we provide an average number of clusters removed *per image*.

7.1.5 Design of Phantom Insert

This section presents a description of the plastic insert that holds the gold fiducial markers and is used in both the RSVP and Quasar phantoms that were introduced in Sections 2.3.1 and 2.3.2. The details of the marker placement inside the insert are necessary for understanding the validation experiments in Sections 7.1.6 and 7.1.7.

The insert contains seven fiducial markers 0.8 mm in diameter and 3 mm long. Such markers are used for daily target position verification for the patients. The markers are placed in a known geometric configuration depicted in Figure 7.3, and the following facts are known about their placement in the phantom insert:

(F1) The *distance* between the centres of the neighbouring markers is 10 mm.

(F2) The centres of all fiducial markers are on *one line*, and that is the axis of symmetry of the cylindrical insert, the direction of which is denoted by a unit vector a in Figure 7.3.

(F3) Axes of symmetry of all markers lie in the *same plane*, which we call a *marker placement plane*.

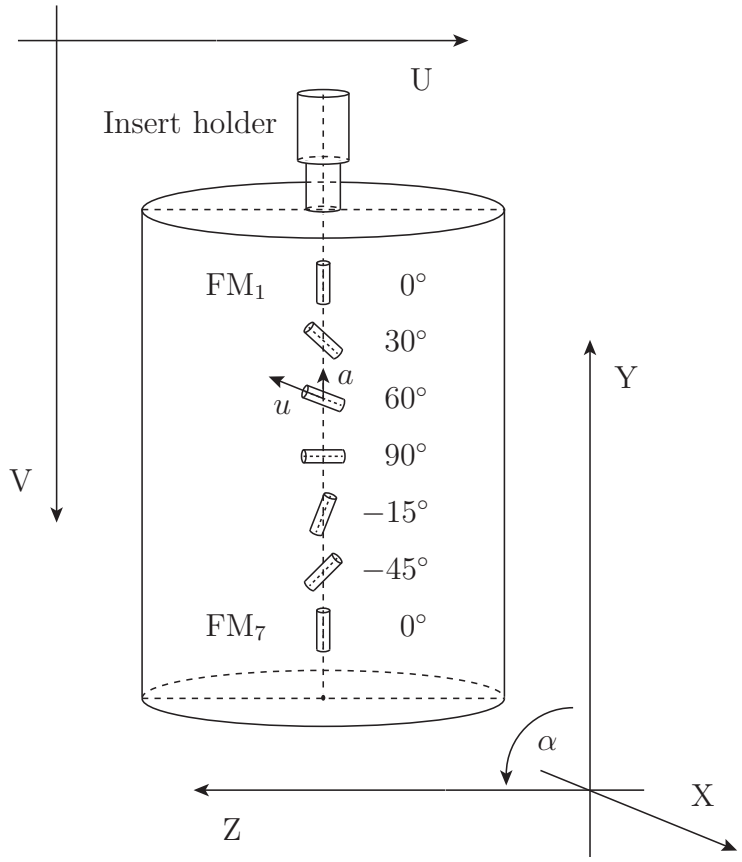


Figure 7.3: Design of the phantom insert with the fiducial markers.

(F4) *Orientation angles* of the fiducial markers are also known. Let u^k be an orientation vector of FM_k, coinciding with the axis of the cylindrical marker. Orientation angles α_k are measured as the angles between a and u^k, where $-90° < \alpha_k \leq 90°$, and are known to be 0, 30, 60, 90, -15, -45, and 0 degrees.

The phantom was imaged in different positions (shifts in the X, Y and Z directions, and rotations around the X axis), however, the angle between a and Y+ (the positive direction of the Y axis) was always smaller than 90°. Also, we made our best effort to place the insert in a way that the marker placement plane is parallel to YZ. This makes the markers appear full-length in the imaging plane when the kV x-ray source is at the X axis (lateral images), and that is how the markers are placed in the prostate patients. We number the fiducial markers starting from the insert holder as $\mathrm{FM}_1, \ldots, \mathrm{FM}_7$. The orientation of the UV coordinate system of the fluoroscopic images shown in Figure 7.3 corresponds to an acquisition setup of left-lateral images, when the imaging x-ray source is located at the X+.

7.1.6 Marker Positional Errors

The validation of the method accuracy requires an existence of some "ground truth" data to compare the results to. In our case, even for the phantom images, it is quite challenging to obtain the exact positions and orientations of the markers in the 3D space. One of the common ways is to compare the results of the method to the marker positions and orientations selected manually in the CT or CBCT images. Here, we take another (automatic) approach: The positional and orientational accuracy of the proposed 3D marker segmentation and modelling procedure is evaluated based on the relative geometry of the markers. In particular, we exploit the phantom insert design facts (F1) – (F4) described in Section 7.1.5.

Based on (F1) and (F2), we compute an *inter-marker distance error* (IMDE), and a *marker positioning error* (MPE), respectively (see Figures 7.4 and 7.5). In this section, we provide formulations and results of the IMDE and MPE for the phantom data sets described in Section 7.1.1.

Inter-Marker Distance Error (IMDE)

Let $c^k_{3\mathrm{D}}$ be a centre of FM_k computed as given in Section 3.2. From the physical phantom design, it is known that the distance between the neighbouring marker centres is $d = 10$ mm. We define the *inter-marker distance error* (IMDE, Figure 7.4) as:

$$\mathrm{IMDE}_k = \|c^k_{3\mathrm{D}} - c^{k+1}_{3\mathrm{D}}\| - d, \quad k = 1, \ldots, n-1, \tag{7.1.1}$$

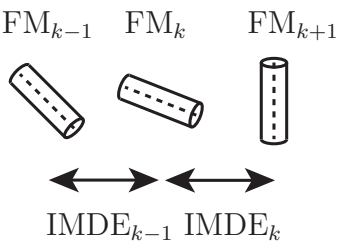


Figure 7.4: Inter-marker distance error (IMDE): A distance in 3D between the marker centres.

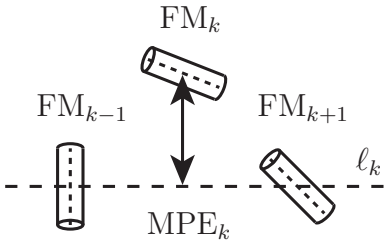


Figure 7.5: Marker positioning error (MPE): A distance in 3D from a marker centre to an estimation of an insert axis, ℓ_k.

Table 7.6: Inter-marker distance error (IMDE).

Data set	IMDE, mm					
	Min	25%	Med	75%	Max	Mean ± Std
F 0.25 mm	−0.08	−0.04	−0.02	0.01	0.05	−0.02 ± 0.04
F 0.5 mm	−0.11	−0.03	0	0.05	0.1	0.01 ± 0.05
F 1 mm	−0.15	−0.04	0.02	0.07	0.2	0.01 ± 0.08
F 2 mm	−0.71	−0.08	0	0.18	0.64	0.01 ± 0.27
F 4 mm	−2	−1.91	−0.76	1.13	2.13	−0.37 ± 1.53
S 0.5 mm	−0.17	−0.21	0.02	0.09	0.17	0.03 ± 0.09
S 1 mm	−0.18	−0.15	−0.03	0.05	0.1	−0.04 ± 0.11

where n is the number of the markers, and $n = 7$ for all phantom data sets used in this section.

As discussed in Section 7.1.4, all markers in the CBCT images of the phantoms were detected correctly, with the exception of the "F 4 mm" data set, where some markers were discarded as they occupied the neighbouring voxels and formed a single large contiguous feature. Also, the first CBCT image of the mentioned data set included one erroneously accepted non-marker feature. We exclude this feature from consideration in this and the following sections.

Figure 7.6 and Table 7.6 present the IMDE results. As expected, the IMDE grows with the increasing voxel size. All data sets show IMDEs that are below 1 mm, except for the "F 4 mm", where the values take the range from −2 to 2.13 mm. In order to place the IMDE in the context of the voxel size, we also present IMDE results as fractions of the corresponding voxel sizes, δ_{vx}: IMDE [vx] = IMDE/δ_{vx} (Figure 7.6, *bottom*). For voxel sizes from 0.25 to 2 mm the maximal and minimal IMDE [vx] values approximately fall within 1/3 of a voxel size, whereas for 4 mm voxels at around 1/2

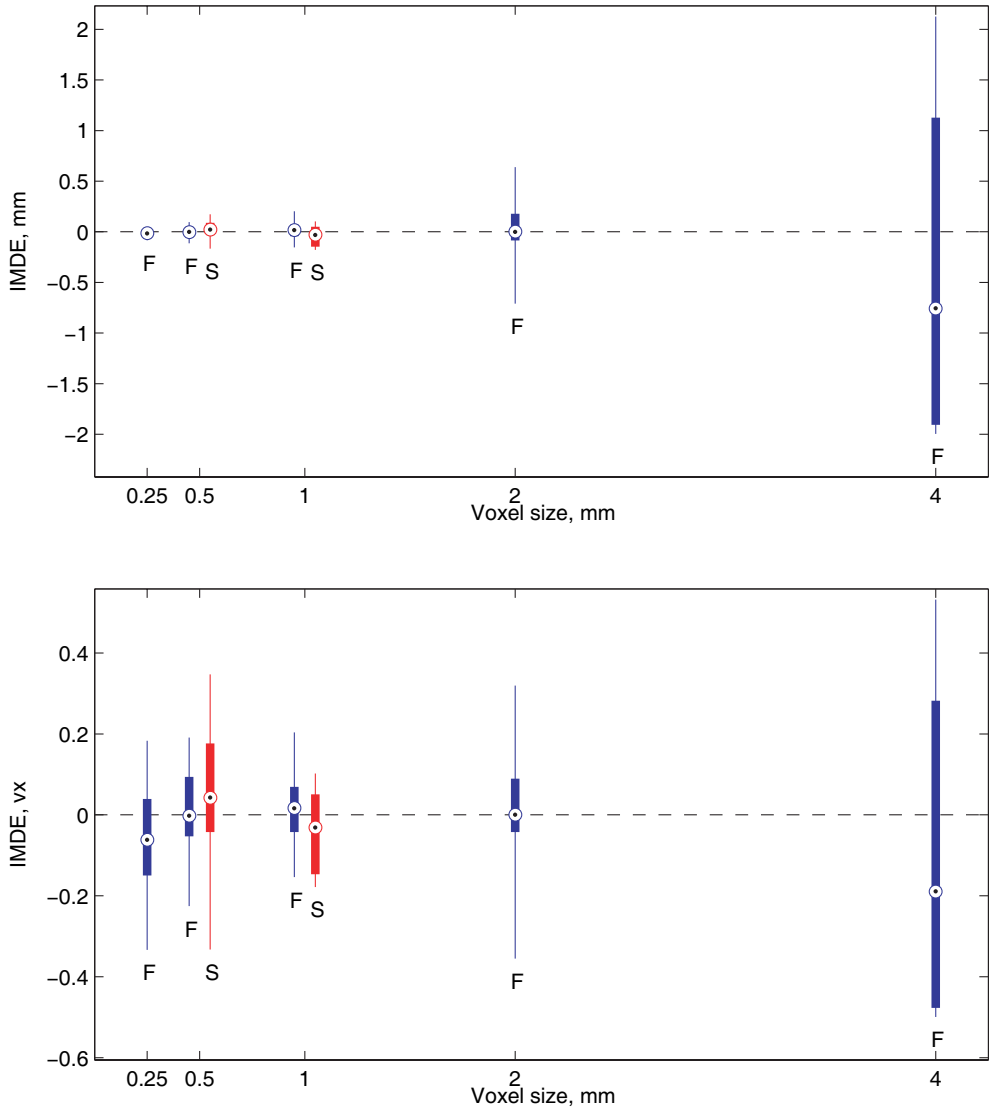


Figure 7.6: Inter-marker distance error (IMDE) results for the phantom data sets in millimeters (*top*) and voxels (*bottom*). The symbols 'F' and 'S' denote `float` and `short` data types, respectively. Table 7.6 provides numerical values for the *top* figure.

with a much larger range of 25^{th} to 75^{th} percentile in comparison to the smaller voxel sizes.

Figure 7.7 shows the IMDE results separately for each inter-marker interval, where interval number k is the interval between FM_k and FM_{k+1}. The Pearson's linear correlation coefficients between the IMDE median values of all data sets except for "F 2 mm" and "S 1 mm" take the range from 0.53 to 0.9, which, in turn, suggests a correlation of the IMDE to the marker orientation in the 3D space.

Marker Positioning Error (MPE)

As described in the design fact (F2) in Section 7.1.5, the phantom insert is manufactured in such a way that all marker centres are positioned on one line, ℓ, and that is the insert's axis of symmetry. We define the *marker positioning error* (MPE) as the distance between the marker centre, c^k_{3D}, and ℓ. However, the precise location of the insert within the XYZ coordinate system is not known, and therefore, we have to estimate ℓ, which is done by the procedure described below.

The estimation of ℓ for each FM_k, ℓ_k, is found by computing the best-fitting line using leave-one-out *principal component analysis* (PCA), i.e., the components are computed based on the centres of all markers except for the FM_k. Let $\Lambda_k = \{c^i_{\mathrm{3D}} \mid i = 1,\ldots,n,\ i \neq k\} \neq \emptyset$ be a set of all marker centres except for FM_k's. Let v^k be the first (largest) principal component of Λ_k. Then,

$$\ell_k = \{\mathrm{mean}(\Lambda_k) + v^k t \mid t \in \mathbb{R}\}. \tag{7.1.2}$$

Let $p^1, p^2 \in \ell_k$, $p^1 \neq p^2$. The MPE for each FM_k, $k = 1,\ldots,n$, is computed as the distance between the point c^k_{3D} and its projection to ℓ_k (see Figure 7.5):

$$\mathrm{MPE}_k = \|c^k_{\mathrm{3D}} - \mathrm{proj}_{\ell_k}(c^k_{\mathrm{3D}})\| = \frac{(p^2 - p^1) \times (c^k_{\mathrm{3D}} - p^1)}{\|p^2 - p^1\|}. \tag{7.1.3}$$

We compute the points p^1 and p^2 by using the largest and the smallest scores of the first principal component in place of t in Equation (7.1.2). The score is the representation of data in the principal component space. Using the minimal and maximal score of the first principal component creates a line segment from p^1 to p^2 that approximately extends from the first to the last marker.

To ensure an accurate MPE value for each marker, the estimates ℓ_k of the insert axis should also be found with sufficient accuracy. For each data set, we test the estimates by analyzing the *inter-line variability* (ILV) by computing the angles between the pairs of the estimates ℓ_i and ℓ_j:

$$\mathrm{ILV}_{ij} = \{\arccos((v^i)^T v^j) \mid i = 1,\ldots,n-1,\ j = i+1,\ldots,n\}. \tag{7.1.4}$$

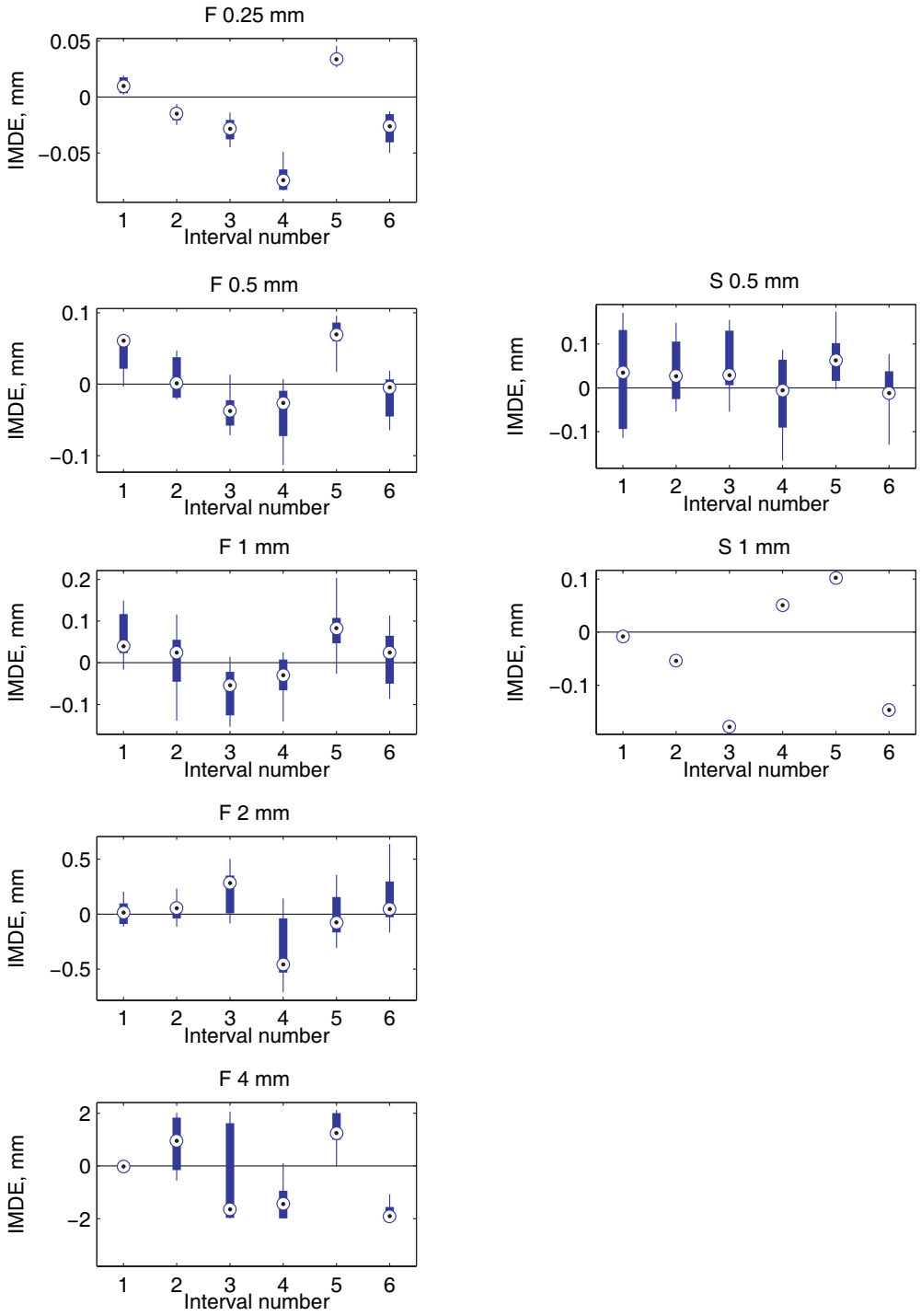


Figure 7.7: IMDE by inter-marker intervals.

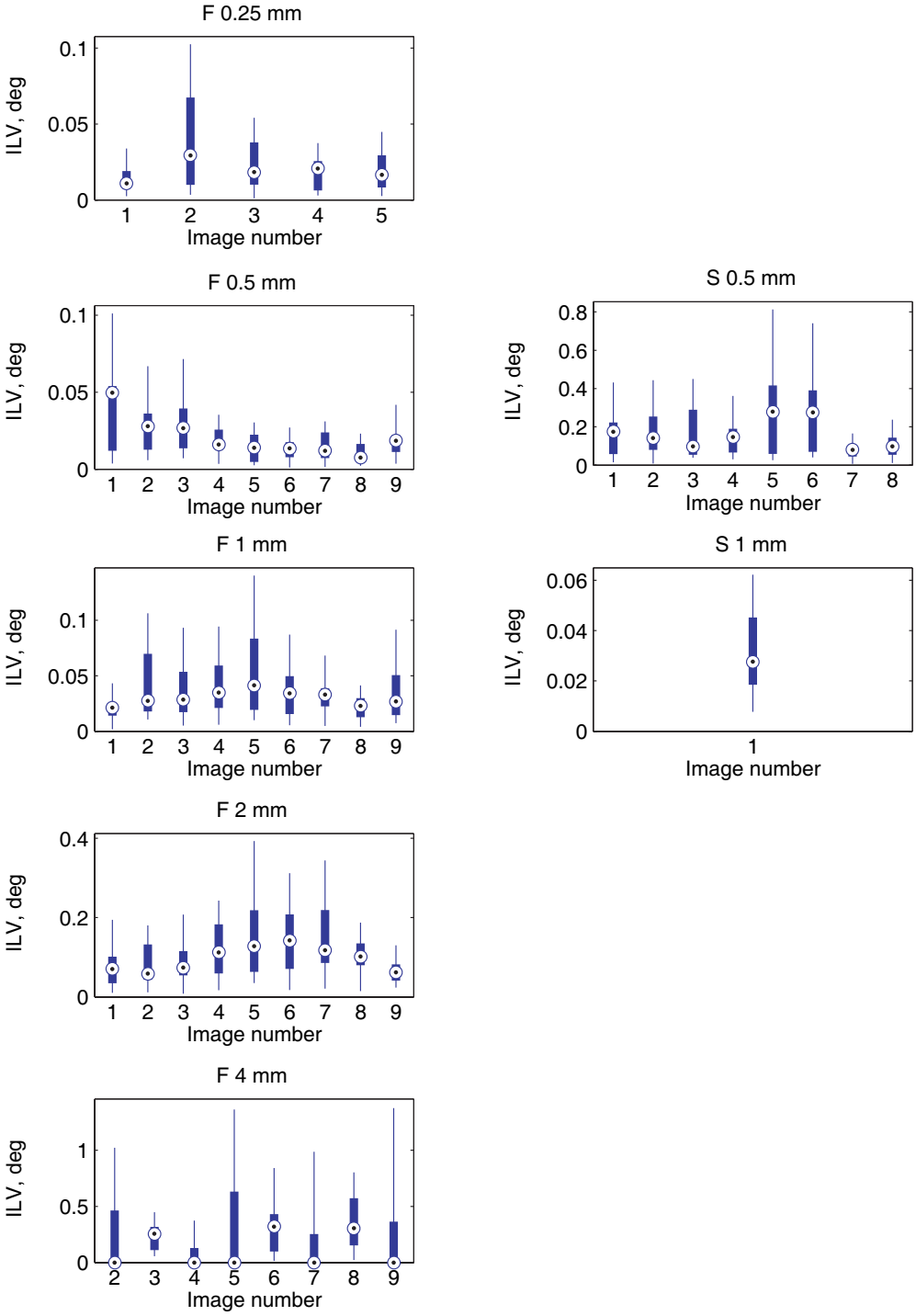


Figure 7.8: Inter-line variability (ILV).

Table 7.7: Marker positioning error (MPE).

Data set	MPE, mm					
	Min	25%	Med	75%	Max	Mean ± Std
F 0.25 mm	0.01	0.02	0.04	0.05	0.11	0.04 ± 0.02
F 0.5 mm	0	0.02	0.03	0.04	0.09	0.03 ± 0.02
F 1 mm	0.01	0.03	0.05	0.07	0.12	0.05 ± 0.03
F 2 mm	0.03	0.09	0.13	0.2	0.35	0.15 ± 0.07
F 4 mm	0.06	0.31	0.45	0.64	1.12	0.48 ± 0.25
S 0.5 mm	0.06	0.26	0.8	1.04	1.36	0.7 ± 0.4
S 1 mm	0.02	0.03	0.04	0.05	0.11	0.05 ± 0.03

The ILV results for the data sets described in Section 7.1.1 are shown in Figure 7.8. As expected, in general the ILV tends to grow with increasing voxel size, with the exception of the "S 0.5 mm" and "S 1 mm" data sets, where the 0.5 mm CBCTs had much larger reconstruction artifacts than the ones with 1 mm voxels. The result for the first image in the "F 4 mm" data set was removed because the presence of the outlier causes almost all line estimates to be erroneous. In order to have a fair base for comparison, we will use the insert axis estimates of the "F 0.5 mm" data set as the ground truth for all data sets with equal or larger voxel size. The position and orientation of the insert axis in 3D were different for the "F 0.25 mm" data set. Therefore, it requires its own insert axis estimates. Their accuracy is comparable to that of the "F 0.5 mm" data set: For both, the maximal ILV is at round 0.1°, which means that the upper bound for displacements between the FM_7's corresponding to different estimates given that the FM_1's coincide, is about 0.1 mm. To compare, for "F 4 mm", with the maximal ILV at around 1.4°, the displacement would be around 1.5 mm, that is, half of the marker size. We conclude that the line estimates found for the "F 0.25 mm" and "F 0.5 mm" data sets in the proposed way are sufficiently accurate and can be used for computation of the MPE.

Figure 7.9 and Table 7.7 present the MPE results. Except for the "S 0.5 mm" data set, the MPE tends to grow with the increasing voxel size. The corresponding MPE results in voxels, i.e., MPE [vx] = MPE/δ_{vx}, shown in Figure 7.9 (*bottom*), suggest a strong dependency between the error and the voxel size. The markers of the "S 0.5 mm" data set include the points of significant reconstruction artifacts that tend to be larger on one side of the markers. An example is shown in Figure 7.10, where the number of the encircled artifact voxels to the left (negative X direction) of the true marker voxels is larger than the number of artifact voxels to the right. This

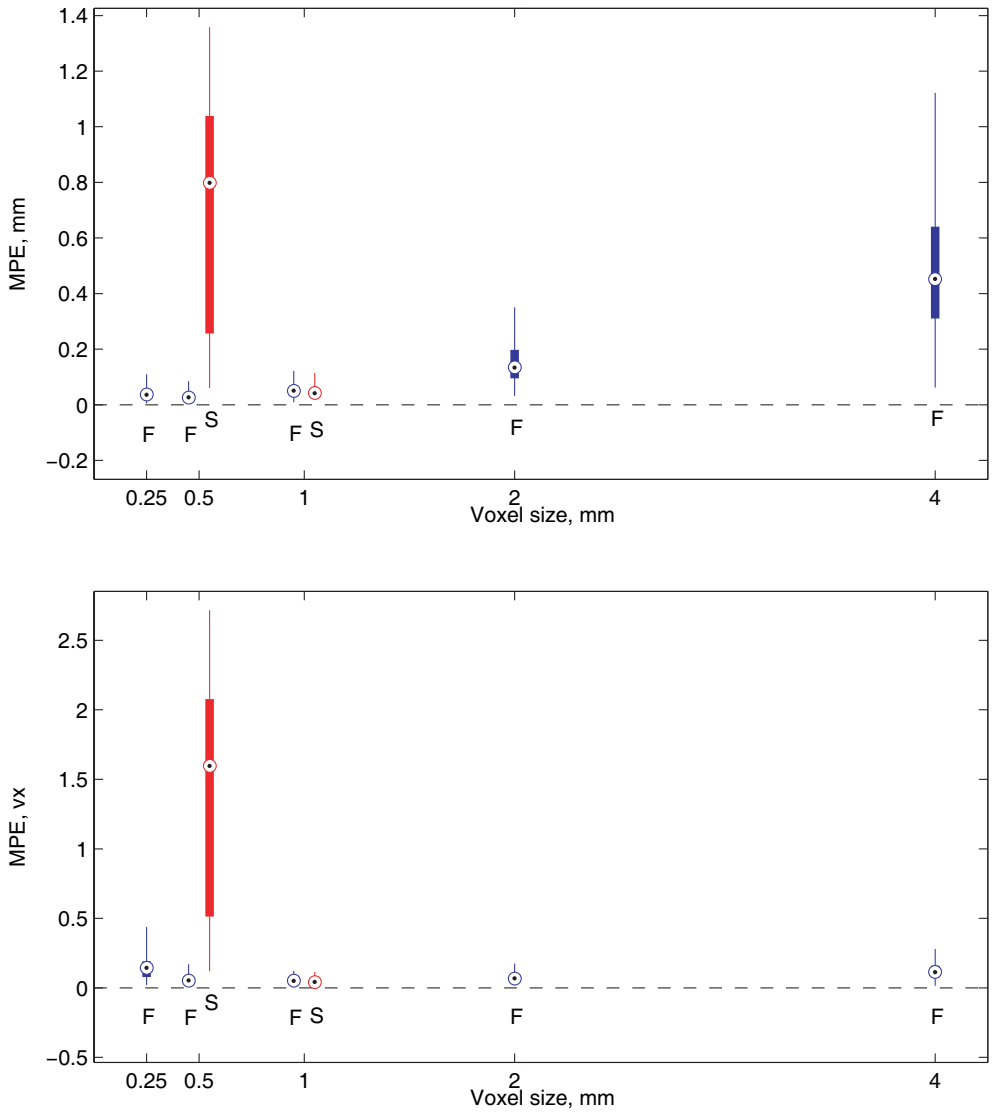


Figure 7.9: Marker positioning error (MPE) for the phantom data sets in millimeters (*top*) and voxels (*bottom*). Table 7.7 provides numerical values for the *top* figure.

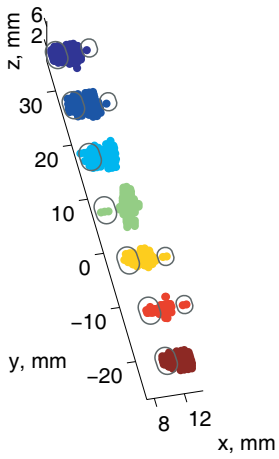


Figure 7.10: In the "S 0.5 mm" data set, it is common for the marker sets to include more artifact voxels (*encircled*) on one side of the true marker voxels, which creates a systematic displacement from the marker set centres to the estimates of the insert axis obtained from the "F 0.5 mm" data set.

creates a significant systematic displacement with a median of around 0.8 mm from all marker centres to the estimates of the insert axis. This displacement can also be viewed in Figure 7.11 (*top right*), where the MPE results are presented separately for each marker. The "by-marker" analysis did not reveal strong correlation between the error and the marker orientation. The only correlation coefficients between the median MPE values that exceeded 0.5 were between "F 0.25 mm" and "F 0.5 mm" (0.86), and between "F 0.5 mm" and "S 1 mm" (0.55).

To conclude, the maximal MPE values for most data sets were well below 1 mm, with the exception of "S 0.5 mm" (1.3 mm) due to reconstruction artifacts, and "F 4 mm" (1.12 mm) due to large voxel size. Based on mostly submillimeter IMDE and MPE results, we conclude that the marker positional accuracy of the proposed 3D segmentation procedure, described in Sections 3.1 – 3.3, is sufficiently high to be successfully used for marker segmentation in the CBCT images in clinical setting. The following section provides validation results for the marker models proposed in Section 3.4. We aim to quantify marker orientational errors to answer the question of whether the marker models can be used in practice.

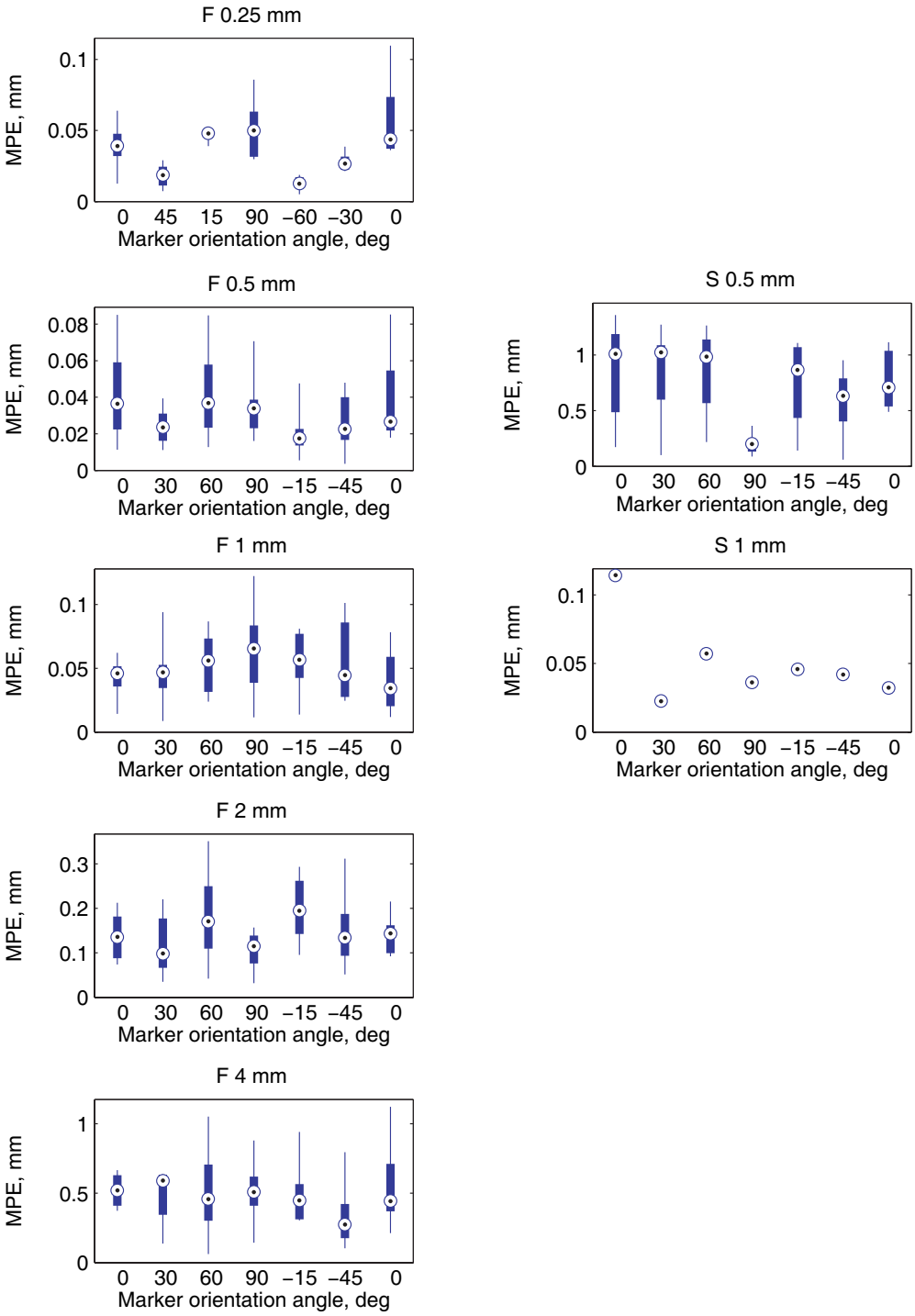


Figure 7.11: MPE by markers.

7.1.7 Marker Orientational Errors

Based on the phantom insert design facts (F3) and (F4) explained in Section 7.1.5, stating that all marker axes lie in one plane, and that their orientations in relation to the phantom insert axis are known, we quantify the accuracy of the marker model orientations by computing the *out-of-plane angle* (OPA) and an *orientation angle error* (OAE, see Figures 7.12 and 7.13, respectively). In this section, we provide formulations and results of the OPA and OAE for the phantom data sets described in Section 7.1.1.

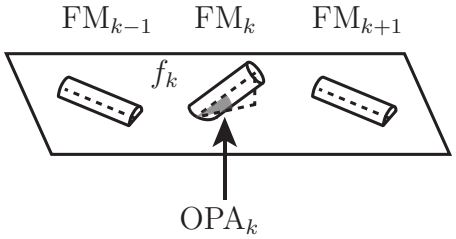


Figure 7.12: Out-of-plane angle (OPA): An angle between a marker model axis and an estimation of a marker placement plane, f_k.

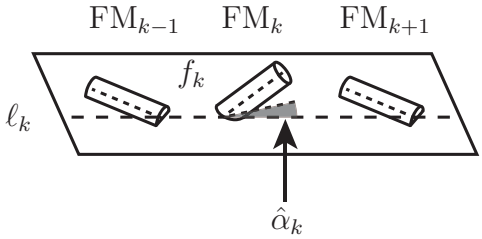


Figure 7.13: Orientation angle error (OAE): A difference between the marker orientation angle, α_k, known from the phantom design, and an angle $\hat{\alpha}_k$ between the projection of the marker model axis to f_k and the estimation of the insert axis, ℓ_k.

Out-of-Plane Angle (OPA)

As described in the phantom design fact (F3) in Section 7.1.5, all marker axes lie in the same plane, which we call the *marker placement plane*, f. Similarly to the MPE, where the exact positioning of the phantom insert axis ℓ was not known and had to be estimated, we also have to estimate f. The estimation is performed in an analogous manner. Let $c^k_{3\mathrm{D}}$ be a centre, and $e^{1,k}$ and $e^{2,k}$ be the endpoints of the marker model corresponding to FM$_k$ found as described in Section 3.4. By $\Phi_k = \{c^i_{3\mathrm{D}}, e^{1,i}, e^{2,i} \mid i = 1, \ldots, n,\ i \neq k\} \neq \emptyset$ we denote a set of all centres and endpoints except for the FM$_k$'s. The application of the PCA to Φ_k gives three component vectors, $v^{1,k}$, $v^{2,k}$, and $v^{3,k}$, where $\|v^{1,k}\| \geq \|v^{2,k}\| \geq \|v^{3,k}\|$, and $v^{1,k} \perp v^{2,k} \perp v^{3,k}$. Then, an estimate of f for FM$_k$, f_k, is the plane spanned by the vectors $v^{1,k}$ and $v^{2,k}$ that is perpendicular to $v^{3,k}$:

$$f_k = \{\mathrm{mean}(\Phi_k) + v^{1,k}t_1 + v^{2,k}t_2 \mid t_1, t_2 \in \mathbb{R}\}. \tag{7.1.5}$$

Let $\hat{u}^k = (e^{2,k} - e^{1,k})/\|e^{2,k} - e^{1,k}\|$ be a marker model orientation vector. Then, the *out-of-plane angle* (OPA) can be computed as:

$$\mathrm{OPA}_k = \left|90^\circ - \arccos\left((\hat{u}^k)^T v^{3,k}\right)\right|. \tag{7.1.6}$$

Similar to the MPE, the OPA can only provide reasonable estimation of marker out-of-plane angles if the estimates for f_k are accurate. For each data set, an *inter-plane variability* (IPV) is quantified by computing the angles between the pairs of plane estimates f_i and f_j:

$$\text{IPV} = \{\arccos\left((v^{3,i})^T v^{3,j}\right) \mid i = 1, \ldots, n-1,\ j = i+1, \ldots, n\}, \tag{7.1.7}$$

where n is a number of the fiducial markers. As can be seen in Figure 7.14, the IPV grows rapidly with the increase of the voxel size. Therefore, we will use the marker placement plane estimates for the "F 0.5 mm" data set for all data sets with the voxel sizes of 0.5 mm and larger. As the "F 0.25 mm" data set was measured with a different phantom, it has its own estimates for f_k that are comparable in accuracy to those of "F 0.5 mm". The maximal IPV is around 1° for both "F 0.25 mm" and "F 0.5 mm", which translates into the distance of approximately 1 mm between the FM_7's in different estimates given that the FM_1's coincide.

We expect that the OPA's should grow with the increase in the voxel size. As was shown in Section 7.1.6, the IMDE and MPE are almost linearly dependent on the voxel size for most data sets (*bottom* graphs in Figures 7.6 and 7.9), i.e., the localization accuracy is inevitably limited by the voxel size. In order to provide a similar context for the OPA results, we define an angle β that corresponds to one voxel size, δ_{vx}.

For the marker models, we believe that a sensible way to define β is depicted in Figure 7.15: We compute a rotation angle that requires to be subtended by the marker model axis, where one endpoint is fixed, and the other travels the distance equal to δ_{vx}. Angle β quantifies the *rotational uncertainty*. For example, if the angular difference between the marker model axis and its estimation is smaller than β, it can not be expected to be accurately recovered. Analogously, we will use β to provide relation between the voxel size and orientation angle error (OAE) presented later in this section. The rotational uncertainty is computed as:

$$\beta = 2 \arcsin\left(\frac{0.5\,\delta_{\text{vx}}}{h}\right), \tag{7.1.8}$$

where $h = 3$ mm is the length of the fiducial marker. The numerical values for β depending on different voxel sizes are given in Table 7.8.

Figure 7.16 and Table 7.8 present the OPA results. We conclude that judging by the out-of-plane angles, the marker orientation can be recovered reliably for the CBCT images with 0.25 and 0.5 mm voxels, and for most images with 1 mm voxels, with the exception being the "S 0.5 mm" data set that had large reconstruction artifacts. While the orientations of most of the marker models in the 1 mm voxel images were

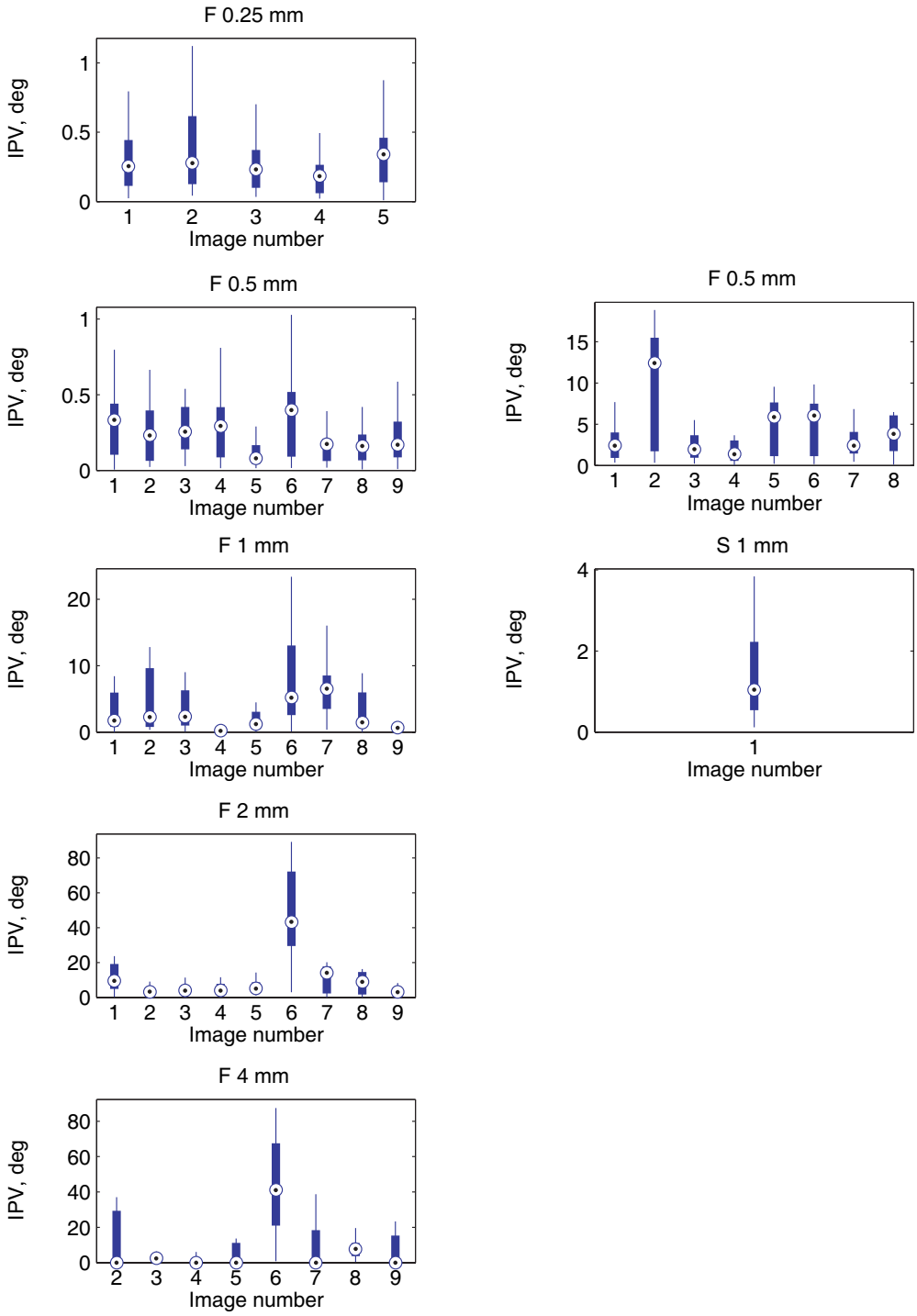


Figure 7.14: Inter-plane variability (IPV).

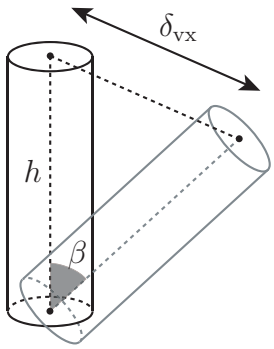


Figure 7.15: Rotational uncertainty is quantified by β, which is a rotation angle that the marker model of length h must execute around one of its endpoints so that the other endpoint travels the distance equal to the voxel size, δ_{vx}. The numerical values for β depending on different voxel sizes are given in Table 7.8.

Table 7.8: Out-of-plane angle (OPA), and rotational uncertainty, β.

Data set	OPA, deg						β, deg/vx
	Min	25%	Med	75%	Max	Mean $\pm$ Std	
F 0.25 mm	0.04	0.55	0.84	1.41	6.76	1.18 ± 1.28	4.78
F 0.5 mm	0.04	0.39	0.7	1.12	4.88	0.91 ± 0.87	9.56
F 1 mm	0.08	0.86	1.79	4.18	81.05	5.84 ± 13.16	19.19
F 2 mm	1.97	7.21	17.82	46.23	86.04	29.47 ± 27.88	38.94
F 4 mm	1.54	10.48	33.1	75.42	87.47	41.62 ± 32.46	83.62
S 0.5 mm	0.07	52.43	76.43	81.37	86.87	61.98 ± 28.85	9.56
S 1 mm	0.36	1.33	4.48	4.88	6.51	3.46 ± 2.27	19.19

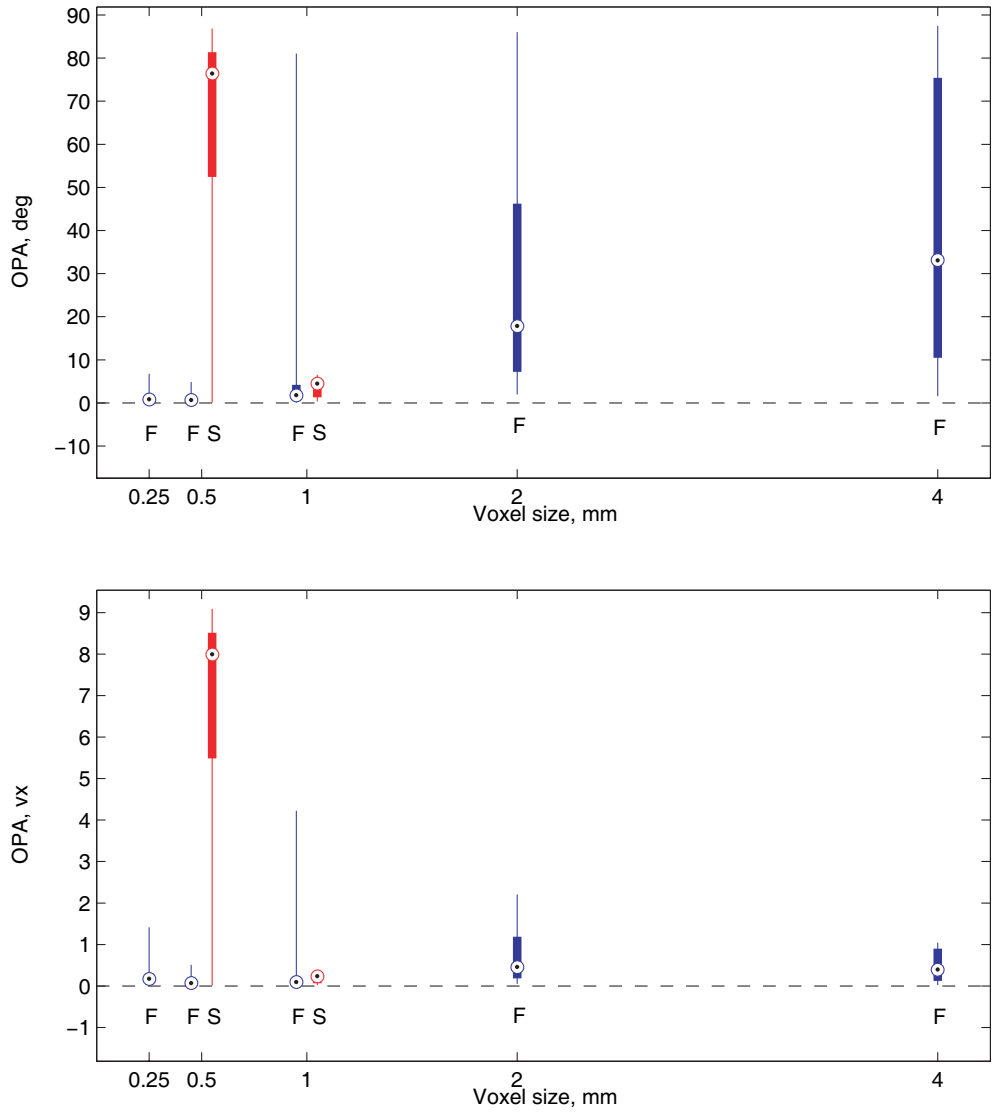


Figure 7.16: OPA for the phantom data sets in degrees (*top*) and scaled by the voxel size (*bottom*) by dividing by β defined in Figure 7.15. Table 7.8 provides numerical values for the *top* figure.

recovered reasonably with medians of 1.79° and 4.48° in "F 1 mm" and "S 1 mm", respectively, some of them went out of plane reaching a value of 81.05°. This means that while the use of the marker models can prove beneficial for our current 1 mm CBCT images, the method should be applied with caution by having error detection procedures in place. For example, if it is known that every effort is made to insert the fiducial markers to appear in one plane, this additional information can be used to either correct the erroneously recovered orientations or detect an unacceptable error and switch to the alternative processing method, such as working with the marker sets directly without reconstructing the models.

The OPA results scaled by the voxel size, OPA [vx] = OPA/β (Figure 7.16, *bottom*), show that, with the exception of "S 0.5 mm" and some of the marker models from the "F 1 mm" data set, the model orientations are recovered with reasonable accuracy. All data sets have their 75^{th} percentile approximately within 1 voxel, and maximal values within 2 voxels. Hence, the ratio between the marker and voxel size plays an important role in the ability to find the marker position and orientation. The "by-marker" analysis, presented in Figure 7.17, does not reveal significant correlation between the OPA's and the marker orientation for most data sets. The only correlation coefficients between the OPA median values that exceeded 0.5 were between "F 0.5 mm" and "F 4 mm" (0.91), and between "F 1 mm" and "F 2 mm" (0.63).

Orientation Angle Error (OAE)

As explained in the design fact (F4) in Section 7.1.5, the orientation angles α_k between the insert axis and the fiducial marker axes are known from the phantom design. These angles are defined in the marker placement plane, where all axes of the fiducial markers should lie according to the design fact (F3). However, as can be understood from the OPA results presented above, the axes of the computed marker models are not necessarily located in the marker placement plane. In this case, it does not make sense to compare α_k to the angles in 3D between the marker models and the phantom axis. In the end, the 3D marker models will be used to generate 2D images that will be compared to lateral fluoroscopic images, i.e., the plane of the 2D projection images will be approximately parallel to the marker placement plane. This means that the observed angles between the projections of the markers to the image plane and the phantom axis can be different from the angles in 3D between the marker models and the phantom axis. Therefore, we propose to compute the *orientation angle error* (OAE) as shown in Figure 7.13: It is the difference between the known orientation angles α_k and the in-plane orientation angles of the marker models, $\hat{\alpha}_k$, found as the angles between the estimation of the phantom axis, ℓ_k, and the projections of the 3D marker model axes to the estimation of the marker placement plane, f_k. As

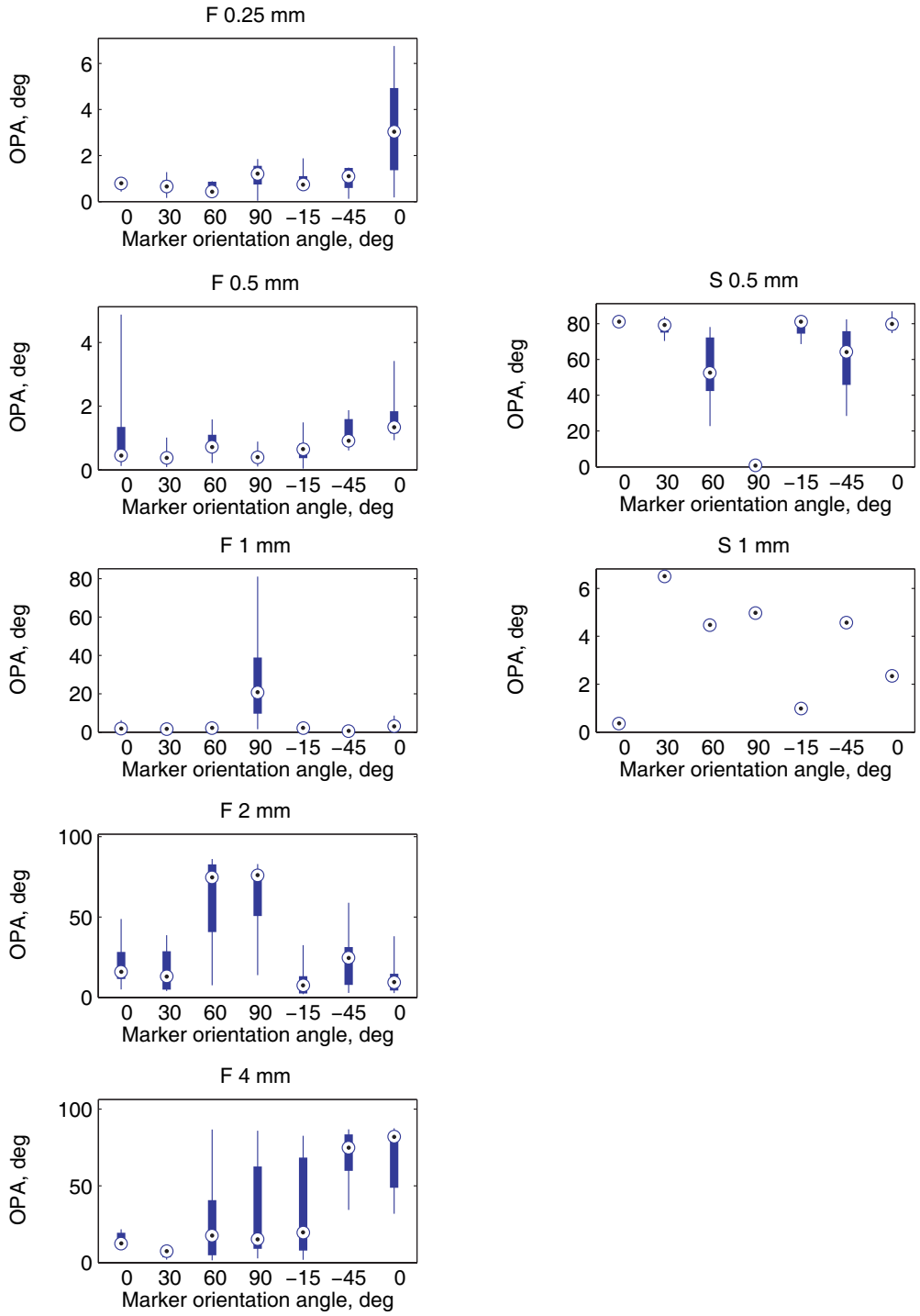


Figure 7.17: OPA by markers.

Table 7.9: Orientation angle error (OAE), and rotational uncertainty, β.

Data set	OAE, deg						β, deg/vx
	Min	25%	Med	75%	Max	Mean $\pm$ Std	
F 0.25 mm	−4.04	−1.63	−0.71	0.09	1.98	-0.7 ± 1.38	4.78
F 0.5 mm	−3.05	−1.82	−1.17	−0.02	3.52	-0.73 ± 1.68	9.56
F 1 mm	−20.8	−11.52	−5.9	8.39	20.15	-2.4 ± 11.06	19.19
F 2 mm	−87.41	−67.72	−25.49	31.8	89.89	-14.47 ± 56.67	38.94
F 4 mm	−87.55	−52.87	−2.64	32.51	77.9	-9.33 ± 51.38	83.62
S 0.5 mm	−89.47	−25.78	10.63	34.33	89.78	8.13 ± 43.72	9.56
S 1 mm	−24.3	−18.76	−3.76	5.59	32.23	-3.22 ± 19.33	19.19

explained in the comments regarding Figure 7.3, the angle between the Y+ axis and the vector a, that is a direction vector of the insert axis, is less than 90°according to our experiment design. If it is greater than 90°, we change the direction of the corresponding vectors $v^{1,k}$ (estimations of vector a for FM_k).

Computation of the projection of the marker model axis to f_k requires finding the projections $e^k_{\mathrm{proj}_1}$ and $e^k_{\mathrm{proj}_2}$ of the model endpoints $e^{1,k}$ and $e^{2,k}$. The distance between $e^{1,k}$ and f_k can be computed as:

$$d_{1,k} = (v^{3,k})^T \left(e^{1,k} - \mathrm{mean}(\Phi_k)\right). \tag{7.1.9}$$

Then, the projection point is:

$$e^k_{\mathrm{proj}_1} = e^{1,k} - d_{1,k}\, v^{3,k}. \tag{7.1.10}$$

The computation for $e^k_{\mathrm{proj}_2}$ is performed analogously. Let $\hat{u}_k = e^k_{\mathrm{proj}_2} - e^k_{\mathrm{proj}_1}$ be the projection of the marker model axes to f_k. Then, the in-plane orientation angle can be computed as:

$$\hat{\alpha}_k = \mathrm{sign}(\hat{u}^k_z) \arccos \frac{(v^{1,k})^T \hat{u}^k}{\|\hat{u}^k\|}. \tag{7.1.11}$$

Finally, we define the OAE as:

$$\mathrm{OAE}_k = \hat{\alpha}_k - \alpha_k. \tag{7.1.12}$$

If $\mathrm{OAE}_k < -90°$ or $\mathrm{OAE}_k > 90°$, the OAE_k is assigned the value of $\mathrm{OAE}_k - \mathrm{sign}(\mathrm{OAE}_k)\, 180°$. As with the OPA's, we use the plane estimates f_k of the "F 0.5 mm" for all data sets, except for the "F 0.25 mm".

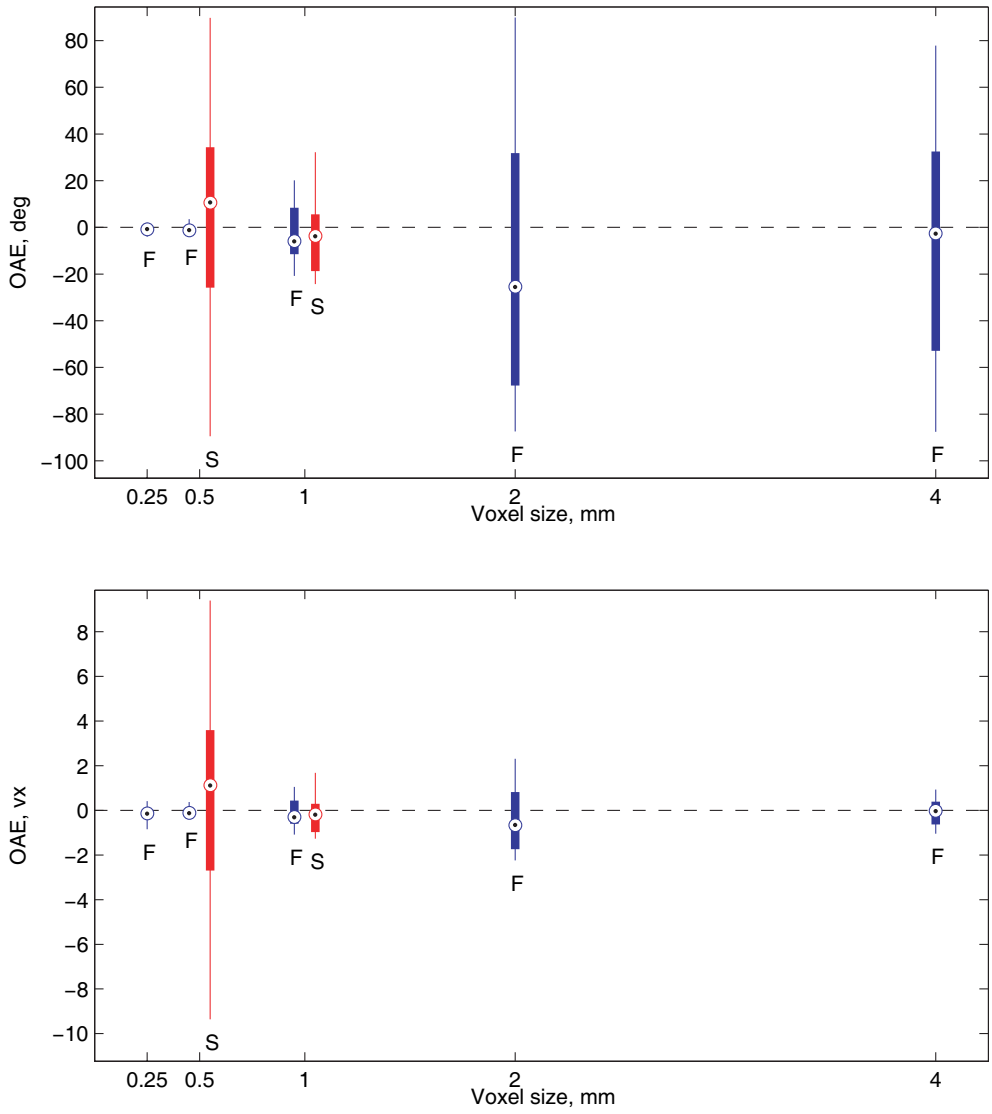


Figure 7.18: OAE for the phantom data sets in degrees (*top*) and scaled by the voxel size (*bottom*) by dividing by β defined in Figure 7.15. Table 7.9 provides numerical values for the *top* figure.

Figure 7.18 and Table 7.9 present the OAE results. In general, the errors grow with the increasing voxel size. Based on the OAE, the marker orientation are well recovered for the CBCT images with 0.25 and 0.5 mm voxels, except the "S 0.5 mm" data set that had large reconstruction artifacts. In the "F 1 mm" data set, the minimal and maximal OAE values are very close in magnitude to the rotational uncertainty, β, and in "S 1 mm" somewhat surpass it. Although the error in orientation of the marker model of around 20° may sound significant, in the CBCTs with 1 mm voxels it corresponds to 1 pixel displacement.

Indeed, the OAE results scaled by the voxel size OAE [vx] = OAE/β (Figure 7.18, *bottom*) show that, with the exception of "S 0.5 mm", the model orientations are reasonably recovered. The "F 2 mm" data set has its minimal and maximal values within 2.3 voxels, while OAE [vx] of all other data sets fall within approximately 1 voxel. The "by-marker" analysis, presented in Figure 7.19, suggests very high correlation between the known marker orientation and the error. All data sets, except for "F 0.25 mm" and "S 0.5 mm", are mutually correlated with correlations of 0.72 to 0.94 between the median OAE values.

Based on the OPA and OAE results, we conclude that the technique of computing the marker models proposed in Section 3.4 can be successfully used for the CBCT images with 0.25 and 0.5 mm voxels, with the exception of the 0.5 mm images reconstructed with the `short` data type, and with some caution for most images with 1 mm voxels. If the voxels are larger than 1 mm, or large reconstruction artifacts are present, with the current hardware and 3D image reconstruction technique, there is no benefit in using the marker models, and the marker localization procedure should proceed by only using the marker sets segmented from the CBCT images as described in Sections 3.1 – 3.3.

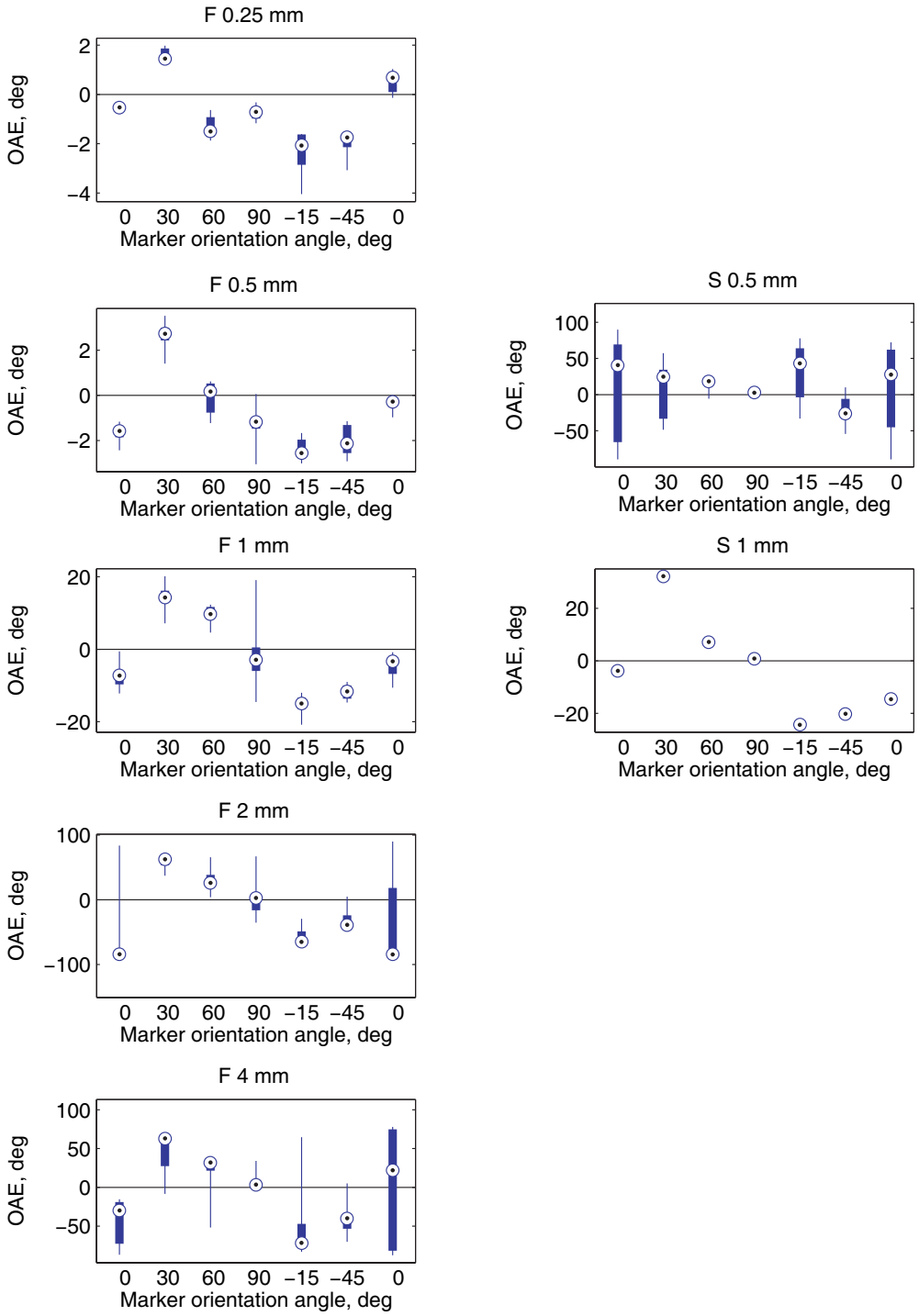


Figure 7.19: OAE by markers.

7.2 MEF Validation

This section discusses the success of the application of the marker enhancement filter (MEF) to phantom and patient images, in particular, in Section 7.2.3. In Section 7.2.4, we provide a comparison of the MEF to the selected existing filters described in Section 4.2.

The success of the marker enhancement depends on characteristics of 2D projection images, such as the noise and contrast, that are affected by the properties of the image acquisition system, the thickness of the volume along the viewing axis, and the density of overlaying structures. In addition, filter performance is expected to be influenced by the choice of the region of interest (ROI) on which filtering is performed: In general, better marker enhancement results are expected on a smaller ROI containing the markers and their immediate background than on a larger ROI with lots of background full of anatomical features. We test the MEF on the data sets described in Section 7.2.2 using several ROIs, the selection of which is discussed in Section 7.2.1.

7.2.1 Selection of Regions of Interest

We used five different ROI sizes for the validation of the MEF. Although filter validation can be performed with no relation to image registration, we prefer to use ROIs that are meaningful in the context of our proposed marker localization and tracking methods. As explained in detail in Section 4.3, the ROIs are computed based on the selection of the geometric transformation and bounds on its parameters, $a^{\max}$. For the 2D rigid transformation, $\mathcal{T}^{2D}$, we compute the ROIs based on the following bounds $a^{\max} = (t_u^{\max}, t_v^{\max}, \gamma^{\max})^T$, where $t_u^{\max}$ and $t_v^{\max}$ are shifts in the U and V directions, and $\gamma^{\max}$ is a bound on the rotation angle:

ROI_1: $a^{\max} = (0 \text{ mm}, 0 \text{ mm}, 0°)^T$;
ROI_2: $a^{\max} = (5 \text{ mm}, 5 \text{ mm}, 0°)^T$;
ROI_3: $a^{\max} = (10 \text{ mm}, 10 \text{ mm}, 3°)^T$;
ROI_4: $a^{\max} = (15 \text{ mm}, 15 \text{ mm}, 5°)^T$;
ROI_5: $a^{\max} = (20 \text{ mm}, 20 \text{ mm}, 5°)^T$.

As explained in Section 4.3, $a^{\max}$ is used to compute the maximal allowable displacements of the markers, ϱ_k, that act as the radii of the circular ROIs Υ_k around the predicted marker positions. The minimal rectangle that includes all Υ_k of the image is Ω_{ROI}. The filter is applied to Ω_{ROI}, as it is easier to operate on a rectangular region that in a straightforward way translates into an array data structure in the software, and then all the values outside of $\bigcup_{k=1}^{n} \Upsilon_k$ are set to zero. Figure 7.20 shows

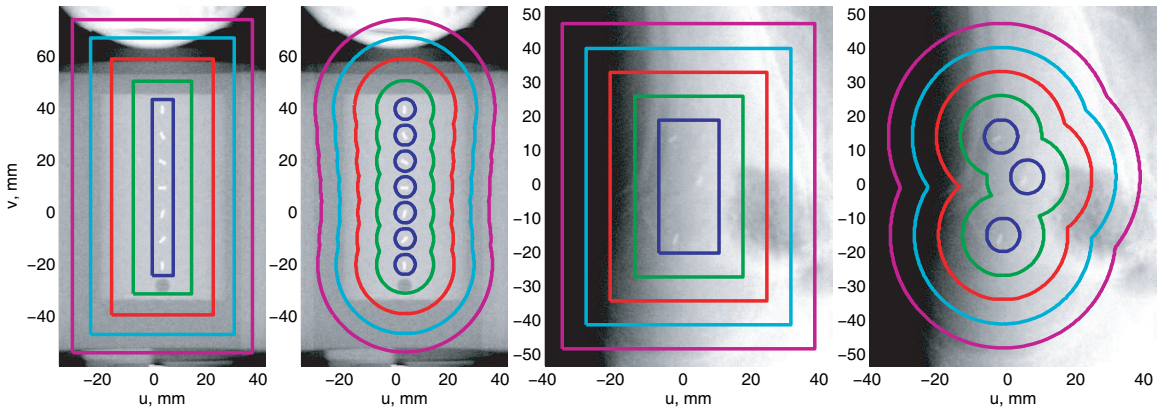


Figure 7.20: ROIs used for the MEF validation, *from left to right*: Rectangular ROI Ω_{ROI}, union of circular ROIs $\bigcup_{k=1}^{n} \Upsilon_k$ for the Quasar phantom, and rectangular and circular ROIs for a prostate patient.

Table 7.10: ROI sizes for the MEF validation.

ROI	ROI sizes			
	Quasar		Patient 1	
	px	mm	px	mm
Ω_{ROI_1}	32×261	8.33×67.97	69×151	17.97×39.32
Ω_{ROI_2}	87×315	22.66×82.03	123×205	32.03×53.39
Ω_{ROI_3}	151×379	39.32×98.7	181×263	47.14×68.49
Ω_{ROI_4}	213×439	55.47×114.32	237×320	61.72×83.33
Ω_{ROI_5}	267×493	69.53×128.39	291×374	75.78×97.4

the respective rectangular and circular ROIs of five sizes for sample fluoroscopic images of the Quasar phantom and a prostate patient. The corresponding sizes[1] of the rectangular ROIs are given in Table 7.10.

7.2.2 Testing Data Sets for MEF Validation

As already mentioned, the MEF is validated and compared to a number of existing filters by using phantom and patient fluoroscopic images. The testing data sets for this section are summarized in Table 7.11 and include the following:

- The RSVP data set consists of nine fluoroscopic image sequences showing the

[1]The sizes in pixels and millimeters are given as guidelines only, as they can vary in different images depending on the respective marker positions in each image.

Table 7.11: Testing data sets for MEF validation.

Data set	Sequences × images	Total number of images
RSVP	9×20	180
Quasar	23×1	23
Patient 1	10×20	200
Patient 2	10×20	200
Patient 3	5×40	200

stationary phantom in different positions and orientations with 20 images in each sequence, resulting in 180 MEF-images.

- The Quasar data set is similar to the "MM 23" set described in the experiments with the temporal filters in Section 4.1.3. The 23 image sequences, each of which is used to produce one MEF-image, show the Quasar phantom exhibiting linear motion at 2.3 mm/sec.
- The patient testing data sets consist of the fluoroscopic sequences of three prostate patients. Patient 1 and 2 had ten fluoroscopic sequences where 20 images were selected, and Patient 3 had five sequences of 40 images, resulting in 200 MEF-images in each patient data set.

The images of both the RSVP and Quasar phantoms contained measurements of seven fiducial markers, while all patients had three. Note that each of the MEF-images is produced for different temporal filter width, ℓ, and different ROI sizes, i.e., for each ℓ and ROI there are 180, 23, or 200 MEF-images in the RSVP, Quasar, and patient data sets, respectively.

7.2.3 MEF Success Rates

In this section, we report the results of the MEF application to five data sets containing fluoroscopic images of the RSVP and Quasar phantoms, and three prostate patients, described in Section 7.2.2. The MEF success rates are reported with respect to different ROI sizes, defined in Section 7.2.1, and the kernel size, ℓ, of the temporal Dolph-Chebyshev filter that was used before the MEF application, as discussed in Section 4.1.3.

Let m be the number of points enhanced by the MEF in one image, which is computed as explained in Section 4.4.2.[2] Let the number of the points enhanced by the MEF in image j that belong to the marker k be denoted by n_{jk}. Then, the

[2]Recall that m is estimated from the knowledge of the marker size, number of the markers, and the pixel size. Therefore, in our case, m is the same for all images of the same data set.

number of true positives for image j can be computed as $\mathrm{TP}_j = \sum_k n_{jk} \leq m$. We define the *success rate* (SR) for each ℓ and ROI as the median across all images in the data set:

$$\text{Median SR} = \text{median}\left\{\frac{\mathrm{TP}_j}{m}\right\} \cdot 100\%. \tag{7.2.13}$$

In addition, we would like to analyze the SRs separately by the markers as the distribution of the MEF-enhanced points can be uneven due to variability in the contrast between the markers in a fluoroscopic image. Let $n_\mathrm{M} = m/n$ be the number of points that are expected to constitute one marker, where n is the number of the markers. We define the *MEF SR of the marker k* in image j as:

$$\text{Marker SR}_{jk} = \frac{n_{jk}}{n_\mathrm{M}} \cdot 100\%. \tag{7.2.14}$$

Since some of the markers can possess better contrast than the others, this number can be greater than 100%.

We designed and implemented an automatic procedure to determine the number of the MEF-enhanced points belonging to each marker, n_{jk}. For each image, the MEF-enhanced points are those whose intensity is larger or equal to $\tau = 0.5$ (Section 4.4.2). In order to determine which of these points belong to particular markers, it is necessary to determine some predicted marker positions, $\hat{c}^{jk}_{\mathrm{2D}} \in \mathbb{R}^2$, and regions $\Phi_{jk} \subset \mathbb{R}^2$ surrounding them that are only large enough to include one marker. For the RSVP data set, where no marker motion was performed, we used a simple projection $\hat{c}^{jk}_{\mathrm{2D}} = \mathcal{P}(c^k_{\mathrm{3D}})$, where $\mathcal{P}$ is the 3D to 2D projection operator defined in Equation (B.5.4), and c^k_{3D} are the 3D marker positions determined from the CBCT image as described in Chapter 3. For the Quasar phantom, the markers of which were performing significant linear motion, the predicted marker positions had to be estimated separately for each kernel size, ℓ, of the temporal Dolph-Chebyshev filter, and we used our 2D marker localization procedure described in Chapter 5 to compute $\hat{c}^{jk}_{\mathrm{2D}}$ (the results of localization were validated visually). For the three patient data sets, where automatic 2D marker localization for small values of ℓ can be challenging, we used the positions $\hat{c}^{jk}_{\mathrm{2D}}$ found by our localization procedure for $\ell = 7$. In this controlled experimental environment, we determined that no displacement larger than $\xi = 2.5$ mm for phantom and 3.5 mm for patient images occurred between $\hat{c}^{jk}_{\mathrm{2D}}$, and the furthest ends of the markers. Therefore, we consider that the points with intensities larger or equal than τ belong to the marker if they are inside the region Φ_{jk} defined as:

$$\Phi_{jk} = \{q \in \mathbb{R}^2 \mid \|q - \hat{c}^{jk}_{\mathrm{2D}}\| \leq \xi\}. \tag{7.2.15}$$

The MEF processing precedes image registration. Therefore, it is important, that the marker enhanced by the MEF is a contiguous feature, where no low intensity points

Table 7.12: Median MEF success rate (SR) for the RSVP data set.

ℓ, frames	Median MEF SR, %				
	ROI_1	ROI_2	ROI_3	ROI_4	ROI_5
1	97.62	87.1	79.76	73.81	67.86
2	100	98.21	92.06	89.68	87.3
3	100	100	96.03	93.45	91.67
4	100	100	97.62	96.23	93.85
5	100	100	98.21	97.22	94.44
6	100	100	99.01	98.21	95.63
7	100	100	99.21	99.21	95.24
8	100	100	100	99.4	97.02
9	100	100	100	99.8	96.63
10	100	100	100	100	97.62
11	100	100	100	100	98.41

separate the high intensity points of the same marker. Hence, n_{jk} is the number of the high intensity points from Φ_{jk} that form the largest contiguous feature.[3] In the MEF, we also increase the radii ϱ_k of Υ_k (Section 4.3) by ξ in order to guarantee the inclusion of all MEF-enhanced points into all ROIs, in particular, the smallest ROI_1.

In the remainder of this section, we discuss the results of MEF application to the testing data sets.

RSVP

The RSVP data set is described in Section 7.2.2. Table 7.12 presents the median MEF SRs defined in Equation (7.2.13) for the MEF images produced from fluoroscopic sequences with five different ROI sizes and Dolph-Chebyshev temporal filtering with $1 \leq \ell \leq 11$. It can be observed that the median MEF SRs increase for higher values of ℓ, with a rapid improvement between $\ell = 1$ and $\ell = 2$. As expected, the median SRs decrease for larger ROIs, where more background and anatomical features tend to be included in the ROI. However, the values remain in the high 90%'s for the most ROIs and ℓ's.

On average, the RSVP images are good approximation to the patient images, as they are noisy and the markers possess relatively low contrast, caused by a relative thickness of the RSVP in the imaging direction. Figure 7.21 shows the distributions of marker SRs defined in Equation (7.2.14). Graphs from *top* to *bottom* correspond

[3] In each feature, the pixels are connected by either their sides or the corners.

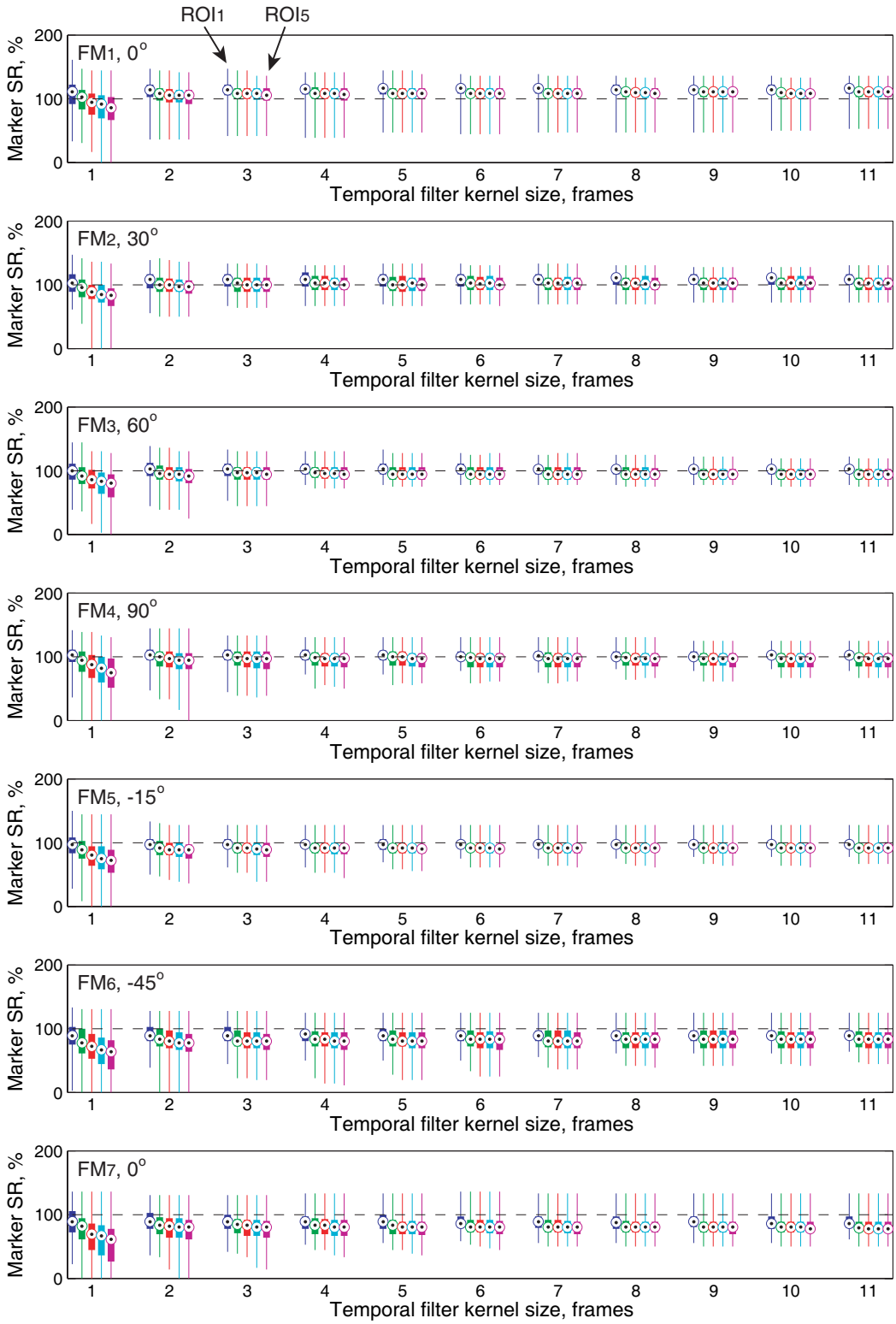


Figure 7.21: Distributions of the marker MEF success rates (SR) for the RSVP data set.

Table 7.13: Median MEF SR for the Quasar data set.

ℓ, frames	Median MEF SR, %				
	ROI_1	ROI_2	ROI_3	ROI_4	ROI_5
1	100	100	100	100	57.54
2	100	100	100	100	60.32
3	100	100	100	100	61.11
4	100	100	100	100	60.32
5	100	100	100	100	60.32
6	100	100	100	100	61.9
7	100	100	100	100	54.37
8	100	100	100	100	57.14
9	100	100	100	100	50.79
10	100	100	100	100	45.63
11	100	100	100	100	40.87

to different fiducial markers, FM_1 to FM_7, schematically shown in Figure 7.3. For each marker and ℓ value, five distributions correspond to marker SR values in ROI_1 to ROI_5, *from left to right*. Due to large noise, the distributions in Figure 7.21 have large ranges, which tend to decrease with increasing ℓ values. It can be noted that the median values of individual marker SRs tend to slightly decrease from FM_1 to FM_7, which is related to the fact that the phantom gradually thickens towards the bottom.

Quasar

The Quasar data set used for the MEF validation is described in Section 7.2.2. Table 7.13 presents the median MEF SR results. The rapid drop in median SRs in ROI_5 is due to the inclusion of the phantom parts that possess significant contrast, and give a strong filter response (the points belonging to the structure at the top in Figure 7.20 (*left*)). For all other ROIs, the higher SR values for the Quasar data set in comparison to the RSVP were expected, as the Quasar images are characterized with a smaller noise and higher marker contrast in comparison to the RSVP.

Figure 7.22 shows the distributions of marker SRs. Except for the ROI_5 mentioned above, there is very small difference between the marker SR values in different ROIs. The linear motion of the phantom insert is performed along the insert axis. It is easy to see that the markers whose axes are perpendicular to the direction of the motion, e.g., FM_4, or at a significant angle, e.g., FM_3, include fewer MEF-enhanced points for higher values of ℓ as they tend to lose contrast in the temporally filtered

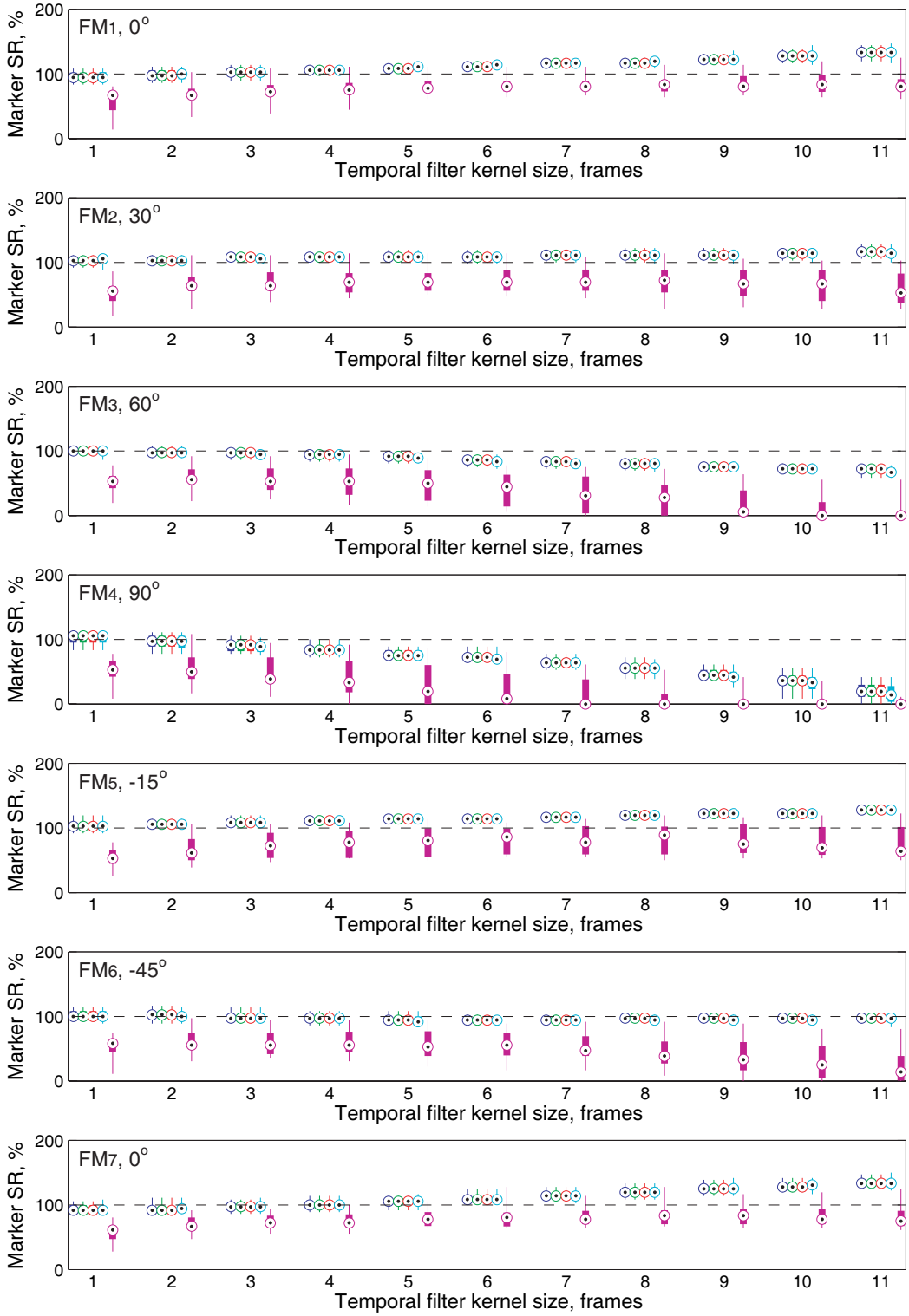


Figure 7.22: Distributions of the marker MEF SR for the Quasar data set.

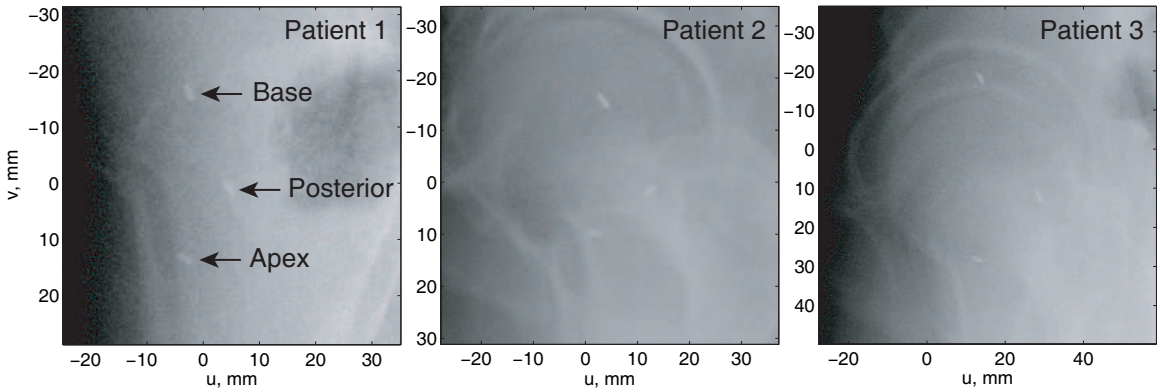


Figure 7.23: Sample fluoroscopic images from the Patient 1 (*left*), Patient 2 (*middle*), and Patient 3 (*right*) data sets with the base, posterior, and apex markers denoted by the *arrows*.

images due to blurring. On the contrary, those whose axes coincide with the motion direction, e.g., FM_1 and FM_7, gain more MEF-enhanced points. For FM_6 that is oriented at 45° to the direction of the motion, the marker SRs are practically the same for any value of ℓ.

Prostate patients

The patient testing data sets are described in Section 7.2.2. Figure 7.23 shows sample fluoroscopic images preprocessed by the Dolph-Chebyshev temporal filter with $\ell = 7$ for each of the patients. The three fiducial markers are typically placed as shown in Figure 7.23 (*left*) and are called the base, posterior, and apex markers. The results for the MEF median SR depending on the choice of ℓ and ROI are presented in Tables 7.14, 7.15 and 7.16 for the Patient 1, 2, and 3 data sets, respectively. The median SRs increase for the higher values of ℓ, with rapid improvement between $\ell = 1$ and $\ell = 2$. As expected, the median SRs decrease for larger ROIs, where more background and anatomical features tend to be included in the ROI.

Separate distributions of the marker SRs for the base, posterior, and apex markers for the Patient 1, 2, and 3 data sets are shown in Figures 7.24, 7.25, and 7.26, respectively. In general, the deviations from the medians, as well as the difference between the marker SRs in different ROIs, tend to decrease with the increasing values of ℓ. The SRs reflect the contrast of the markers in the fluoroscopic images that can be visually assessed in Figure 7.23. Often, out of the three markers, the measurements of the base marker tend to possess the highest contrast while the posterior marker typically has the lowest contrast.

Table 7.14: Median MEF SR for the Patient 1 data set.

ℓ, frames	Median MEF SR, %				
	ROI_1	ROI_2	ROI_3	ROI_4	ROI_5
1	92.59	83.33	76.85	69.44	52.78
2	100	97.22	92.59	89.81	70.37
3	100	100	97.22	95.37	75.46
4	100	100	99.07	96.3	76.85
5	100	100	100	98.15	78.7
6	100	100	100	99.07	78.24
7	100	100	100	99.54	78.7
8	100	100	100	100	80.56
9	100	100	100	100	81.48
10	100	100	100	100	83.33
11	100	100	100	100	83.33

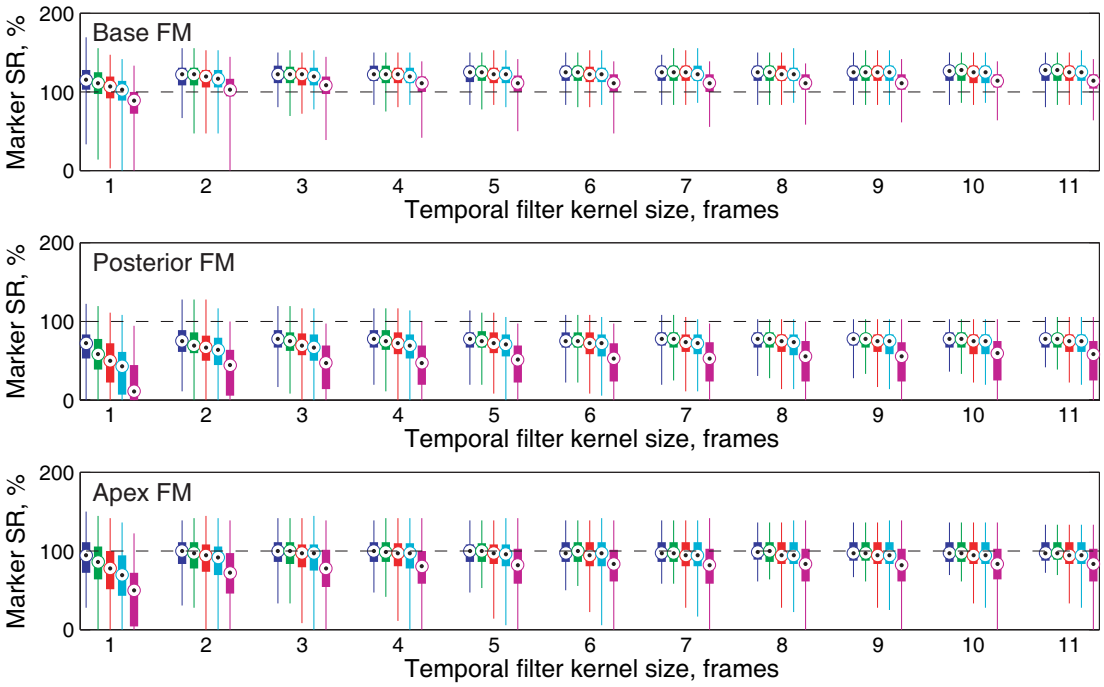


Figure 7.24: Distributions of the marker MEF SR for the Patient 1 data set.

Table 7.15: Median MEF SR for the Patient 2 data set.

ℓ, frames	Median MEF SR, %				
	ROI_1	ROI_2	ROI_3	ROI_4	ROI_5
1	100	98.61	92.59	88.89	75
2	100	100	99.07	94.44	78.24
3	100	100	100	98.15	79.63
4	100	100	100	98.15	81.48
5	100	100	100	100	80.56
6	100	100	100	100	81.02
7	100	100	100	100	81.48
8	100	100	100	100	81.94
9	100	100	100	100	81.48
10	100	100	100	100	82.41
11	100	100	100	100	81.94

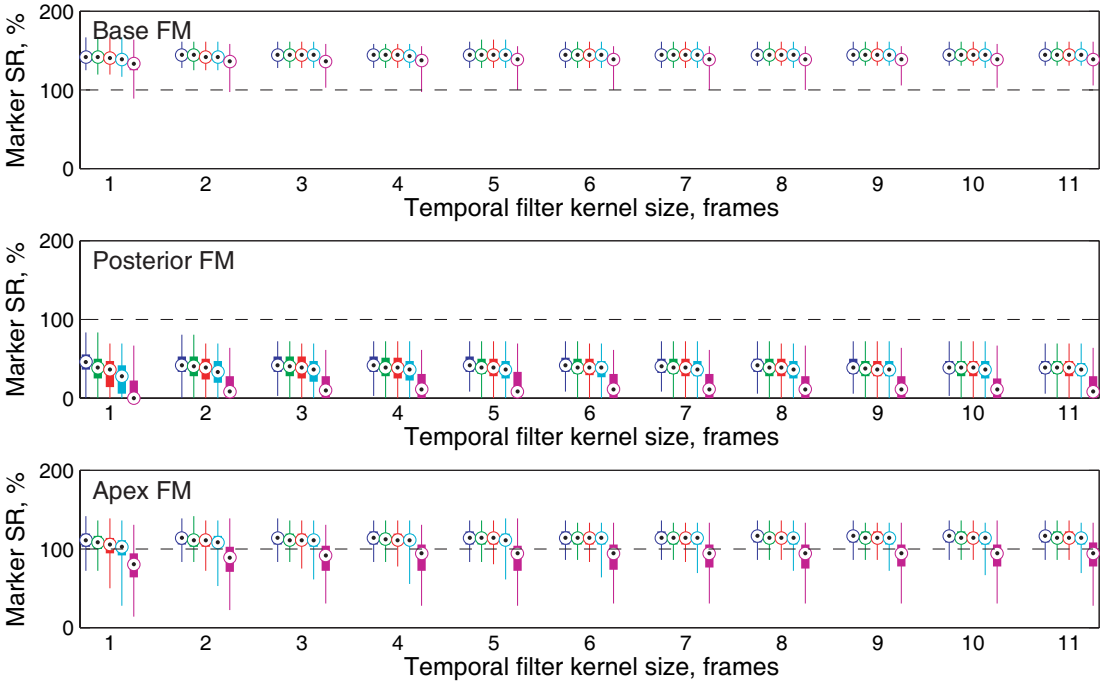


Figure 7.25: Distributions of the marker MEF SR for the Patient 2 data set.

Table 7.16: Median MEF SR for the Patient 3 data set.

ℓ, frames	Median MEF SR, %				
	ROI_1	ROI_2	ROI_3	ROI_4	ROI_5
1	98.15	88.89	73.15	63.89	55.09
2	100	98.15	86.11	75.93	68.52
3	100	100	88.89	80.56	71.3
4	100	100	89.81	83.33	74.07
5	100	100	90.74	84.72	76.39
6	100	100	90.74	85.19	76.85
7	100	100	90.74	87.04	78.7
8	100	100	91.67	87.04	80.56
9	100	100	91.67	88.43	81.48
10	100	100	93.06	88.43	81.48
11	100	100	93.52	88.89	83.8

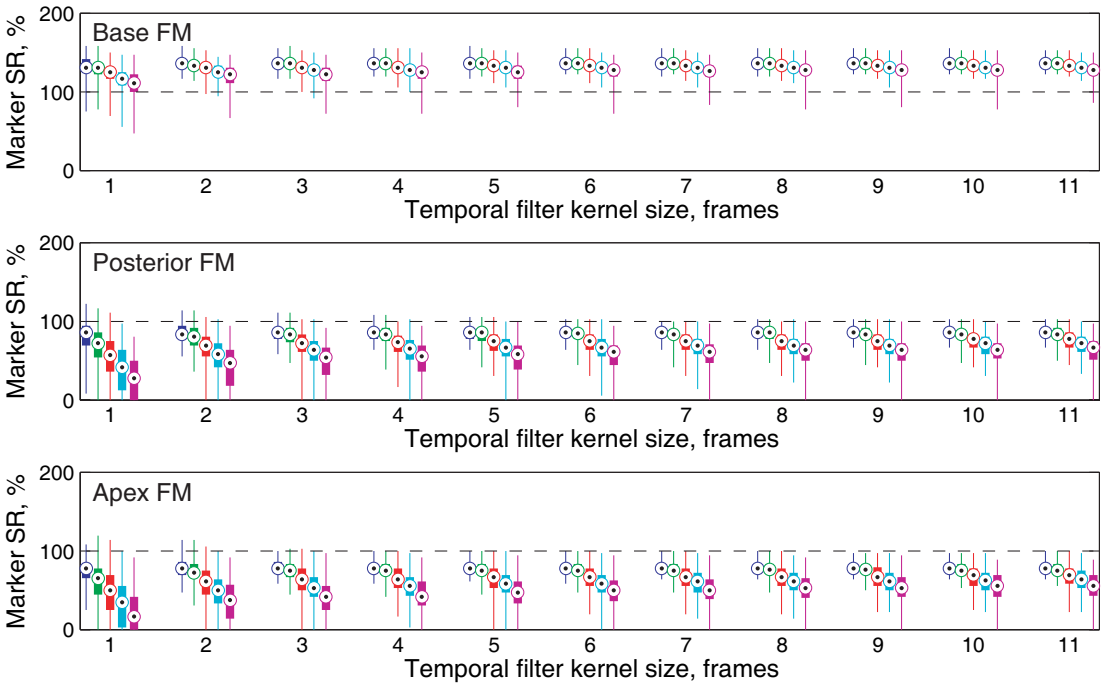


Figure 7.26: Distributions of the marker MEF SR for the Patient 3 data set.

Summary

While the numbers are different for different data sets, ROI sizes and ℓ, it was important to validate the MEF, and analyze the resulting trends. The assessment of the SRs for the RSPV, Quasar phantom, and prostate patient data sets suggests that, for most values of ℓ and ROI sizes, the MEF succeeds in enhancing the marker points, and that this points constitute around $90-100\%$ of all the points enhanced by the MEF in an image. As explained in Section 7.2.1, the choice of the ROI depends on the bounds on the parameters of the geometric transformation used in the localization and tracking procedures, and as such, may be separately defined for each particular experiment depending on the assumptions about the motion characteristics. The best choice of ℓ is subject to successful 2D marker localization, and, as such, is discussed in Section 5.4.

7.2.4 Comparing MEF to Other Filters

As explained in Sections 4.2 and 4.4.1, the existing filters either did not provide sufficient marker enhancement or tended to distort marker shapes. The MEF successfully enhances marker points, as was demonstrated in Section 7.2.3, and preserves the shapes, as was explained in Section 4.4.1 and verified visually in multiple processed images. To complete the validation, this section provides a numerical comparison between the MEF and several existing filters. Taking into consideration that multiple methods require an application of a binary threshold to the filtered image for successful marker detection, e.g., [35, 169, 159], we believe that a reasonable way to compare the filters is to present the results in the form of *receiver operating characteristic (ROC) curves* [55]. In general, the ROC curve illustrates the performance of a binary classifier system as its discrimination threshold is varied. In filter comparison, it means that a series of intensity thresholds is applied to the filtered images. The points whose intensity is higher than the threshold value are split into the true and false positives, TP and FP, where TP are the points that belong to the markers, and FP to the background. Then, the number of TP is plotted versus the number of FP for different threshold values. Often, instead of TP and FP, the graph shows the TP and FP rates, TPR and FPR, which are computed as the fraction of TP out of the positives and the fraction of FP out of the negatives, respectively. The higher the TP or TPR and the lower the FP or FPR, the better is the performance of a filter.

The MEF is compared to the following filters: MEK [160], LoG^2 [127], Sato's blob filter [189], and Frangi's line filter [63] defined in Equations (4.2.9), (4.2.15) (LoG), (4.2.20), and (4.2.27), respectively, and the simple step thresholding $\mathcal{C}_{\text{step}}$ defined in Equation (4.4.35). These filters were selected based on visual assessment

as the ones providing the best marker enhancement, except for the global thresholding, which is used as a baseline. Due to the fact that the filters can produce images in different intensity ranges, the thresholds are selected in relation to the intensity range as will be explained in more detail later in this section. The evaluation procedure for each of the filters proceeds as follows:

- Circular and rectangular ROIs, Υ_k and Ω_{ROI}, are computed as explained in Section 4.3 in a discrete fluoroscopic image temporally preprocessed by the Dolph-Chebyshev filter;
- A filter is applied to Ω_{ROI}, resulting in image $\mathtt{A}$;
- Intensity values outside of Υ_k are set to the minimum intensity found in the image;
- A sequence of intensity thresholds is selected. We compute the threshold interval as $\Delta\theta = (\max\{\mathtt{A}\} - \min\{\mathtt{A}\})/(n_{\text{pt}} + 1)$, where $n_{\text{pt}} = 20$ is the number of thresholds. Then,

$$\theta_i = \min\{\mathtt{A}\} + i\Delta\theta, \quad i = 1, \ldots, n_{\text{pt}}. \tag{7.2.16}$$

Let $\check{p}$ be the grid points, $\check{p} \in \Gamma^2 \cap \Omega_{\text{ROI}}$. The following steps are performed for each threshold θ_i:

- All points $\{\check{p} \notin \bigcup_k \Upsilon_k \mid \mathtt{A}[\check{p}] \geq \theta_i\}$ are counted towards FP;
- The points $\{\check{p} \in \Upsilon_k \mid \mathtt{A}[\check{p}] \geq \theta_i\}$ are processed according to the procedure described in Section 7.2.3 for the computation of the marker MEF SR defined in Equation (7.2.14): The points within each region[4] Φ_{jk} defined in Equation (7.2.15) are split into contiguous features and the points of the largest feature are counted towards TP, while all other points are counted towards FP;
- If the number of TP for any Φ_{jk} is larger than the estimated number of points that should constitute a marker, n_{M}, used in Equation (7.2.14), then the number of FP is increased by $(\text{TP} - n_{\text{M}})$, and the number of TP for this particular Φ_{jk} is set to n_{M}.

As mentioned above, the ROC curve often shows the fraction of TP out of the positives, i.e., the true positive rate (TPR), or sensitivity, versus the fraction of false positives out of the negatives, i.e., the false positive rate (FPR), or one minus the specificity. The expected number of the marker points, i.e., the number of *positives*, P, is known: It can be computed based on the number and size of markers as explained in Section 4.1.2. Hence, the TPR = TP/P. The number of *negatives*, N, is the number of pixels in the image minus P, and FPR = FP/N. However, for our filter

[4] Υ_k is formed based on $a^{\max}$, and Φ_{jk} is computed in the experimental setting as a way of selecting marker points within Υ_k based on geometry and intensity.

evaluation, the comparison of TPR to FPR does not make sense, as N $\gg$ P: For example, if TP = FP, the result of the filtering can not be considered acceptable for marker detection or image registration purposes, as there are as many enhanced points belonging to the markers as to the other features. Nonetheless, in this example FPR $\ll$ TPR, which in a general case would suggest good performance. Therefore, for filter evaluation, it is better to compare TP to FP rather than TPR to FPR. However, to place TP and FP in the context of the ideal filter performance, we scale them by P:

$$\widetilde{\mathrm{TP}}_j(\theta_i) = \frac{\mathrm{TP}_j(\theta_i)}{\mathrm{P}} \cdot 100\%, \tag{7.2.17}$$

$$\widetilde{\mathrm{FP}}_j(\theta_i) = \frac{\mathrm{FP}_j(\theta_i)}{\mathrm{P}} \cdot 100\%, \tag{7.2.18}$$

where TP_j and FP_j are the TP and FP numbers for the image j. To be consistent with the results in Section 7.2.3, we report the percentage rather than the fraction. For each threshold value θ_i, we report median values of $\widetilde{\mathrm{TP}}_j$ and $\widetilde{\mathrm{FP}}_j$ across all processed images. Due to the fact that both TP and FP are scaled by P, $0\% \le \widetilde{\mathrm{TP}}_j \le 100\%$, while the values of $\widetilde{\mathrm{FP}}_j$ can be much larger than 100%.

MEF benchmarking on RSVP data set

The filter performance is evaluated using the RSVP data set that consists of nine fluoroscopic image sequences, 38 images each, totalling in 342 images. We present the results for the images preprocessed by the Dolph-Chebyshev temporal filter with kernel sizes $\ell = 7$ and $\ell = 1$ (no temporal filtering). The filtering is performed over ROI_3 defined in Section 7.2.1 as it is the one that includes a sufficient amount of the background but does not have too many irrelevant measurements, such as those of the phantom structures.

The $\widetilde{\mathrm{TP}}$ and $\widetilde{\mathrm{FP}}$ results in the form of the ROC curves for the MEF and existing filters, such as LoG^2, Frangi's line, Sato's blob, MEK, and a simple threshold technique as a baseline are presented in Figure 7.27. The figures on the *left* show $0\% \le \widetilde{\mathrm{FP}} \le 100\%$ and $0\% \le \widetilde{\mathrm{TP}} \le 100\%$, and the figures on the *right* give a zoom-in into $0\% \le \widetilde{\mathrm{FP}} \le 50\%$, $50\% \le \widetilde{\mathrm{TP}} \le 100\%$ for $\ell = 1$ (*top figures*) and $\ell = 7$ (*bottom figures*). The corresponding numerical values at different thresholds θ_i are listed in Tables 7.17 – 7.20 for the existing filters, and in Tables 7.21 and 7.22 for the MEF. The *underlined row* in the tables approximately corresponds to the middle value in the intensity range. For the MEF, this is the value of the threshold that separates the candidate marker points from the background.

Without preliminary temporal filtering ($\ell = 1$), the MEF clearly provides the best performance of the considered filters by having the smallest $\widetilde{\mathrm{FP}}$ and the highest

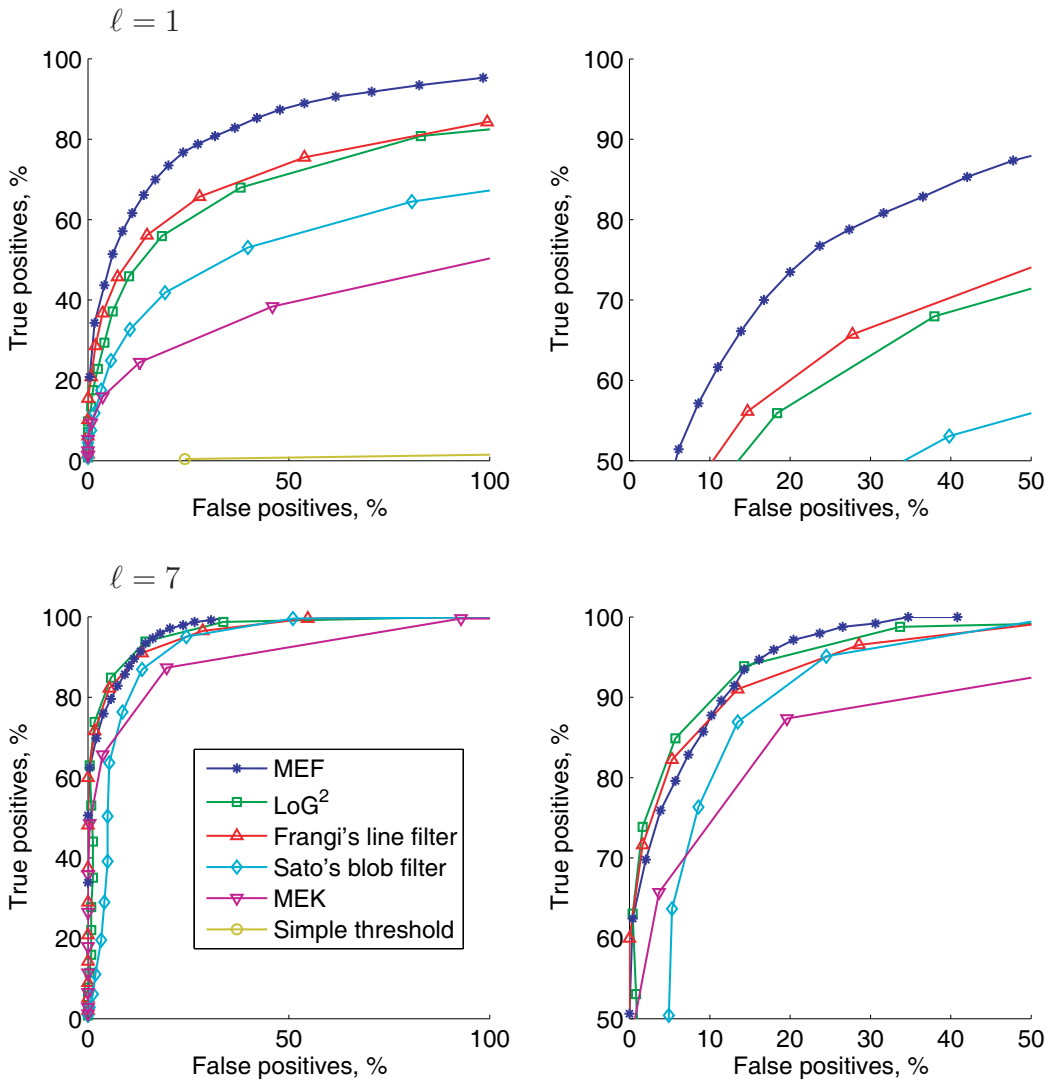


Figure 7.27: Receiver operating characteristic (ROC) curves ($\widetilde{\mathrm{TP}}$ versus $\widetilde{\mathrm{FP}}$) for the MEF and existing filters for $\ell = 1$ (*top left*) and $\ell = 7$ (*bottom left*). The figures on the *right* provide zoom-in into the corresponding graphs on the *left*.

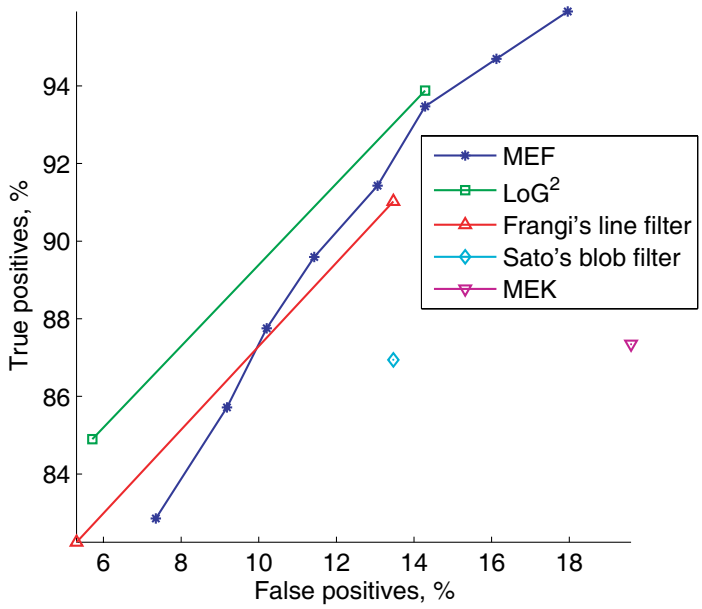


Figure 7.28: ROC curves for the MEF and existing filters for $\widetilde{\text{TP}} \geq 80$, $\widetilde{\text{FP}} \leq 20$, and $\ell = 7$.

$\widetilde{\text{TP}}$ values, as can be seen from Figure 7.27 (*top*). For $\ell = 7$ (Figure 7.27, *bottom*), other filters, especially LoG^2 and Frangi's line filter, tend to have similar $\widetilde{\text{TP}}$ and $\widetilde{\text{FP}}$ values. Often, the best achievable $\widetilde{\text{TP}}$ values come at a cost of unreasonably high $\widetilde{\text{FP}}$ values. For example, the value of $\widetilde{\text{TP}} = 100\%$ for $\ell = 7$ can be achieved for the LoG^2 at a minimal value of $\widetilde{\text{FP}} = 91.02\%$, for the Frangi's line filter at 101.43%, for the Sato's blob filter at 120.82%, for the MEK at 350.61%, for the simple threshold at 10,164.49%, and for the MEF at just 34.69%. The MEF also provides a larger range of thresholds that correspond to acceptable values of $\widetilde{\text{TP}}$ and $\widetilde{\text{FP}}$. For example, see Figure 7.28 that shows the points for $\ell = 7$, for which $\widetilde{\text{TP}} \geq 80\%$ and $\widetilde{\text{FP}} \leq 20\%$.

By evaluating the performance of the MEF on several data sets reported in Section 7.2.3 and comparing its performance to a number of other filters in this section, we conclude that the MEF is an excellent choice for the marker enhancement task, and, to the best of our knowledge, is the best among the available filters.

7.2.5 MEF Variations

In this section, we report the results on the MEF variations, that may be more suitable for some applications. The results for the MEF, MEF without application of the contrast enhancement function (CEF, Section 4.4.2) as the final MEF step, the magnitude-only filter (MOF, Section 4.4.5), and the MOF without final CEF are

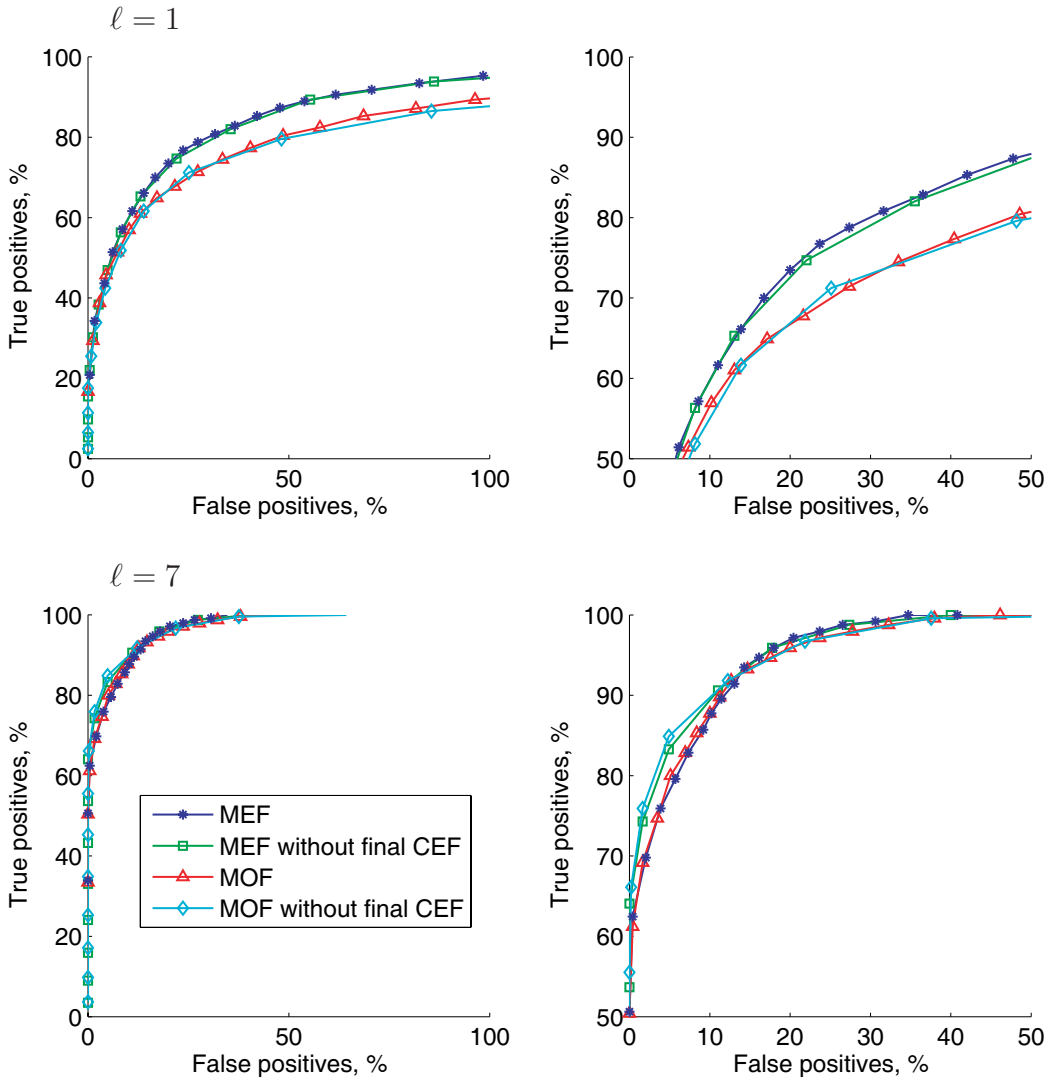


Figure 7.29: ROC curves ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the variations of the MEF for $\ell = 1$ (*top left*) and $\ell = 7$ (*bottom left*). The figures on the *right* provide zoom-in into the corresponding graphs on the *left*.

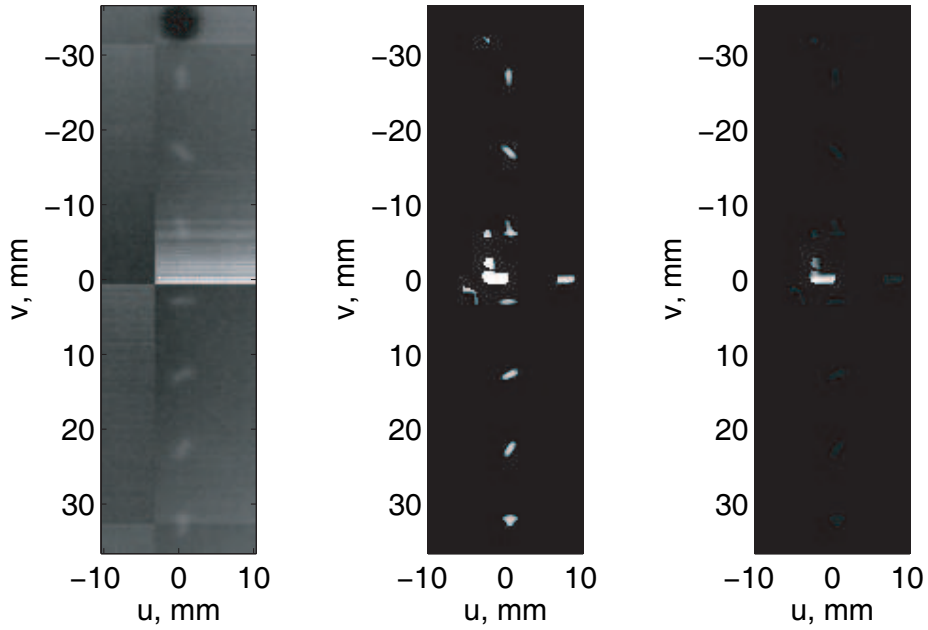


Figure 7.30: A portal image of the RSVP (*left*), and the results of the MEF application with (*centre*) or without (*right*) CEF as the final MEF step.

presented in Figure 7.29, and Tables 7.21 and 7.22. For the RSVP data set described in Section 7.2.4, all four variations provided fairly similar results, hence all can be used in marker enhancement.

As discussed in Section 4.4.5, we used the full MEF in the 2D marker localization procedure, as the expected marker displacement between the CBCT and fluoroscopic images can be significant, and we used the MOF in tracking to reduce computational time. The application of the CEF as the final step of the MEF or MOF may or may not be necessary. As an example, consider a portal image shown in Figure 7.30 (*left*). It contains artifacts related to the portal flat-panel detector, where the background consists of rectangular regions of different and varying intensities, which creates high intensity features at the region borders. These features can give very high filter response, much higher than the markers, as can be seen in Figure 7.30 (*centre*). Since the feature is not very large, the application of the CEF as the final MEF step amplifies marker points as well. On the contrary, without the CEF, only the artifact feature is amplified (Figure 7.30, *right*). A different example would be a fluoroscopic image of a prostate patient, where one of the markers may be overlayed by a high-intensity feature, such as the bones. In this case, other markers tend to attract more MEF points, and at the final CEF step the first marker can be further suppressed. In this case, it may be desirable to omit the final CEF application.

Table 7.17: Receiver operating characteristic (ROC) values ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the LoG^2, Frangi's line, and Sato's blob filters for $\ell = 1$.

i in θ_i	LoG^2		Frangi's line		Sato's blob	
	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)
1	100	(3730.61)	100	(4647.14)	100	(6018.98)
2	100	(1269.59)	100	(2854.69)	100	(4973.27)
3	98.37	(475.51)	100	(1664.29)	100	(3775.1)
4	91.43	(191.22)	100	(961.63)	100	(2668.78)
5	80.82	(82.86)	98.57	(549.18)	98.78	(1750.61)
6	67.96	(37.96)	95.51	(314.08)	95.92	(1067.96)
7	55.92	(18.37)	90.61	(177.14)	93.06	(608.37)
8	45.92	(10.2)	84.29	(99.39)	86.12	(325.51)
9	37.14	(6.12)	75.51	(53.88)	76.33	(163.06)
10	29.39	(4.08)	65.71	(27.76)	64.49	(80.61)
11	22.86	(2.45)	56.12	(14.69)	53.06	(39.8)
12	17.55	(1.22)	45.71	(7.35)	41.84	(19.18)
13	13.47	(0.82)	36.73	(3.67)	32.65	(10.41)
14	9.8	(0)	28.57	(1.84)	24.9	(5.71)
15	7.14	(0)	20.82	(0.82)	17.55	(3.27)
16	4.9	(0)	15.51	(0)	11.84	(1.63)
17	3.27	(0)	10.2	(0)	7.55	(0.82)
18	2.24	(0)	6.12	(0)	4.49	(0)
19	1.63	(0)	3.27	(0)	2.45	(0)
20	0.82	(0)	1.63	(0)	0.82	(0)

Table 7.18: ROC values ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the LoG^2, Frangi's line, and Sato's blob filters for $\ell = 7$.

i in θ_i	LoG^2		Frangi's line		Sato's blob	
	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)
1	100	(1511.43)	100	(3332.65)	100	(7526.53)
2	100	(309.8)	100	(1273.27)	100	(5211.02)
3	100	(91.02)	100	(627.14)	100	(3036.12)
4	98.78	(33.67)	100	(337.55)	100	(1522.04)
5	93.88	(14.29)	100	(184.9)	100	(680.61)
6	84.9	(5.71)	100	(101.43)	100	(287.14)
7	73.88	(1.63)	99.59	(54.69)	100	(120.82)
8	63.06	(0.41)	96.53	(28.57)	99.59	(51.02)
9	53.06	(0.82)	91.02	(13.47)	95.1	(24.49)
10	44.08	(1.22)	82.24	(5.31)	86.94	(13.47)
11	35.1	(1.22)	71.63	(1.63)	76.33	(8.57)
12	27.76	(0.82)	60	(0)	63.67	(5.31)
13	22.04	(0.82)	48.16	(0)	50.41	(4.9)
14	15.92	(0.82)	37.55	(0)	39.18	(4.9)
15	11.43	(0.41)	28.98	(0)	28.98	(4.08)
16	7.76	(0.2)	20.82	(0)	19.59	(3.27)
17	5.31	(0)	14.29	(0)	11.02	(1.84)
18	3.27	(0)	8.98	(0)	6.12	(1.22)
19	1.63	(0)	5.1	(0)	2.86	(0.41)
20	0.82	(0)	2.04	(0)	0.82	(0)

Table 7.19: ROC values ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the MEK and simple threshold for $\ell = 1$.

i in θ_i	MEK		Threshold	
	$\widetilde{\text{TP}}$	$(\widetilde{\text{FP}})$	$\widetilde{\text{TP}}$	$(\widetilde{\text{FP}})$
1	100	(20 201.63)	100	(24 454.29)
2	100	(20 187.76)	100	(24 450)
3	100	(20 083.88)	100	(24 422.65)
4	100	(19 808.78)	100	(24 351.63)
5	100	(18 816.12)	100	(24 231.02)
6	100	(16 571.63)	100	(24 043.06)
7	100	(12 937.35)	100	(23 744.08)
8	100	(8713.27)	100	(23 361.22)
9	100	(5090.2)	100	(22 883.47)
10	100	(2534.29)	100	(22 311.84)
11	97.96	(1100.41)	100	(21 641.84)
12	84.08	(412.86)	100	(20 852.86)
13	59.18	(140)	100	(19 705.92)
14	38.37	(45.92)	100	(17 953.06)
15	24.49	(12.86)	100	(15 210.61)
16	15.92	(3.67)	100	(11 105.92)
17	9.39	(0.82)	94.29	(5952.04)
18	5.31	(0)	42.45	(2149.18)
19	2.45	(0)	5.71	(397.14)
20	1.22	(0)	0.41	(24.08)

Table 7.20: ROC values ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the MEK and simple threshold for $\ell = 7$.

i in θ_i	MEK		Threshold	
	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)
1	100	(20 202.04)	100	(24 453.47)
2	100	(20 167.14)	100	(24 432.65)
3	100	(20 010.82)	100	(24 366.12)
4	100	(19 072.45)	100	(24 266.94)
5	100	(15 535.31)	100	(24 131.02)
6	100	(9126.94)	100	(23 917.35)
7	100	(3780)	100	(23 627.55)
8	100	(1200.61)	100	(23 256.94)
9	100	(350.61)	100	(22 808.78)
10	99.59	(92.86)	100	(22 285.1)
11	87.35	(19.59)	100	(21 741.43)
12	65.71	(3.67)	100	(21 125.92)
13	48.57	(0.41)	100	(20 293.88)
14	35.92	(0)	100	(19 083.06)
15	26.53	(0)	100	(17 325.31)
16	17.96	(0)	100	(14 619.18)
17	11.43	(0)	100	(10 164.49)
18	6.53	(0)	93.06	(5051.63)
19	2.86	(0)	38.78	(1701.84)
20	1.22	(0)	3.27	(125.71)

Table 7.21: ROC values ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the MEF variations for $\ell = 1$.

i in θ_i	MEF		MEF no CEF		MOF		MOF no CEF	
	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)
1	95.31	(98.37)	100	(906.53)	95.1	(197.14)	100	(5049.39)
2	93.47	(82.45)	100	(564.29)	93.06	(148.16)	100	(3439.8)
3	91.84	(70.61)	100	(381.43)	91.02	(118.57)	100	(2182.24)
4	90.61	(61.63)	100	(267.76)	89.39	(96.33)	100	(1357.76)
5	88.98	(53.88)	98.78	(186.73)	87.14	(81.63)	100	(814.08)
6	87.35	(47.76)	96.73	(128.16)	85.31	(68.57)	98.37	(480)
7	85.31	(42.04)	93.88	(86.12)	82.45	(57.76)	96.33	(279.8)
8	82.86	(36.53)	89.39	(55.31)	80.41	(48.57)	92.65	(157.96)
9	80.82	(31.63)	82.04	(35.51)	77.35	(40.41)	86.53	(85.51)
10	78.78	(27.35)	74.69	(22.04)	74.49	(33.47)	79.59	(48.16)
11	76.73	(23.67)	65.31	(13.06)	71.43	(27.35)	71.22	(25.1)
12	73.47	(20)	56.33	(8.16)	67.76	(21.63)	61.63	(13.88)
13	70.00	(16.73)	46.94	(4.90)	64.90	(17.14)	51.84	(8.16)
14	66.12	(13.88)	38.37	(2.65)	61.02	(13.06)	42.45	(4.29)
15	61.63	(11.02)	30.2	(1.22)	56.94	(10.2)	33.88	(2.04)
16	57.14	(8.57)	22.04	(0.41)	51.43	(7.35)	25.51	(0.82)
17	51.43	(6.12)	15.51	(0)	45.71	(4.49)	17.55	(0)
18	43.67	(4.08)	9.8	(0)	38.78	(2.86)	11.43	(0)
19	34.29	(1.63)	5.31	(0)	29.39	(1.22)	6.53	(0)
20	20.82	(0.41)	2.45	(0)	16.73	(0)	2.45	(0)

Table 7.22: ROC values ($\widetilde{\text{TP}}$ versus $\widetilde{\text{FP}}$) for the MEF variations for $\ell = 7$.

i in θ_i	MEF		MEF no CEF		MOF		MOF no CEF	
	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)	$\widetilde{\text{TP}}$	($\widetilde{\text{FP}}$)
1	100	(40.82)	100	(285.92)	100	(60.61)	100	(4111.22)
2	100	(34.69)	100	(173.67)	100	(46.12)	100	(1649.39)
3	99.18	(30.61)	100	(119.18)	99.59	(37.96)	100	(771.43)
4	98.78	(26.53)	100	(89.8)	98.78	(32.24)	100	(392.65)
5	97.96	(23.67)	100	(68.57)	97.96	(27.76)	100	(203.47)
6	97.14	(20.41)	100	(53.06)	97.14	(23.67)	100	(111.63)
7	95.92	(17.96)	100	(40)	95.92	(20)	100	(64.29)
8	94.69	(16.12)	98.78	(27.35)	94.69	(17.55)	99.59	(37.55)
9	93.47	(14.29)	95.92	(17.76)	93.27	(14.69)	96.73	(21.84)
10	91.43	(13.06)	90.61	(11.02)	91.84	(12.65)	91.84	(12.24)
11	89.59	(11.43)	83.27	(4.9)	89.8	(11.22)	84.9	(4.9)
12	87.76	(10.2)	74.29	(1.63)	87.76	(10)	75.92	(1.63)
13	85.71	(9.18)	64.08	(0)	85.31	(8.37)	66.12	(0.2)
14	82.86	(7.35)	53.67	(0)	82.86	(6.94)	55.51	(0)
15	79.59	(5.71)	43.27	(0)	80	(5.1)	45.31	(0)
16	75.92	(3.88)	33.06	(0)	74.69	(3.47)	34.9	(0)
17	69.8	(2.04)	24.08	(0)	69.18	(1.63)	25.31	(0)
18	62.45	(0.41)	15.92	(0)	61.22	(0.41)	17.14	(0)
19	50.61	(0)	8.98	(0)	50.41	(0)	9.8	(0)
20	33.88	(0)	3.47	(0)	33.47	(0)	3.67	(0)

Table 7.23: Testing data sets for validation of 2D marker localization.

Data set	Measure-ments of	Number of CBCTs and fluoroscopic sequences	Number of images per sequence	Total number of fluoroscopic images
P1	Patient 1	10	≈ 160	1618
P2	Patient 2	10	≈ 160	1619
P3	Patient 3	5	≈ 160	820
RSVP	RSVP	9	47	423
P1-S	Patient 1	10	5	50
P2-S	Patient 2	10	5	50
P3-S	Patient 3	5	10	50
RSVP-S	RSVP	9	5	45

7.3 Marker Localization in 2D Space

7.3.1 Testing Data Sets for Marker Localization

The components of the 2D localization procedure are evaluated on the testing data sets described in Table 7.23. The data sets P1, P2, P3, and RSVP contain fluoroscopic sequences and their corresponding CBCT images, one CBCT image per sequence. Each fluoroscopic sequence consists of around 160 images for the patient data sets, and 47 for the phantom.

For some experiments, we also use the subsets of the P1, P2, P3 and RSVP data sets, denoted by P1-S, P2-S, P3-S, RSVP-S, respectively. They are formed so that ten nonconsecutive fluoroscopic images are selected from each sequence of the P3 data set, and five from each sequence of the P1, P2, and RSVP data sets, resulting in 50 images (and registrations) for the patient data sets and 45 for the RSVP.

The accuracy of the 2D marker localization is validated on the RSVP-S data set in Section 7.3.2. This data set is also used to justify the choice of the dimensionality of the geometric transformation in Section 7.3.3. The full data sets, P1, P2, P3, and RSVP, are used to quantify the computational cost and success rates of the proposed image generation methods in Section 7.3.4. The reduced data sets, P1-S, P2-S, P3-S, and RSVP-S, are employed in selecting the best strategy for an optimization starting point in Section 7.3.5, and to justify the selection of the temporal filter width in Section 7.3.6.

7.3.2 Accuracy of 2D Marker Localization

To estimate the accuracy of a localization method, it is common to compare it to other methods, the accuracy for which has been already established, by applying the method to the same data. The positions or the displacements of the sought-for features may be known from the "ground truth", which can be obtained from the construction of the phantom, other phantom images and motion or positioning geometry [43, 59, 166], from synthetically constructed images [72], or manual selection of the markers [78, 142, 159]. To improve the accuracy, manual localization is usually performed several times on the same images, and then these multiple results are used to estimate the "ground truth" data [78, 142].

Due to the fact that in the current clinical setup the precise calibration parameters for the fluoroscopic images were not available (this issue is addressed in detail in Appendix B.4), and had to be estimated from either a flexmap (calibration file) or from calibration parameters of the CBCT projections of the corresponding CBCT image, the CBCT image can not be used directly to infer the 2D reference marker positions in the fluoroscopic images. Therefore, to estimate the accuracy of our 2D marker localization procedure, we compared it to manual localization. We have used the RSVP-S data set described in Section 7.3.1. Five nonconsecutive frames were selected from each of the nine fluoroscopic sequence, totalling 45 images. As we believe that the manual selection of an edge of the marker is easier and can be more precise than selecting its centre,[5] the 2D marker positions in each image were computed as a middle point between the two endpoints selected by an operator. Although the RSVP was placed in different positions in each of the fluoroscopic sequences and their respective CBCT images, it always remained stationary during the acquisition of the sequence and the CBCT. Therefore, the *reference marker positions* for each sequence were computed as the averages of the positions in the five selected frames of each sequence.

The experiment for estimation of accuracy of our 2D marker localization is summarized in Figure 7.31. Five frames were selected in each of the nine fluoroscopic sequences $\mathcal{F}^0, \ldots, \mathcal{F}^8$. As a result of manual 2D localization, we obtained reference marker positions $(q^k)^i$ for each sequence $\mathcal{F}^i$, where $k = 1, \ldots, n$, and $n = 7$ is the number of the fiducial markers. Then, we formed nine data sets, where data set j contained a CBCT image $\mathcal{V}^j$ and the same five images from each of the fluoroscopic sequences $\mathcal{F}^0, \ldots, \mathcal{F}^8$ that were used in manual localization (Figure 7.31, *right*). The 2D-3D image registration between $\mathcal{V}^j$ and $\mathcal{F}_f^i$ was performed, resulting in transformation parameters $\hat{a}^{jif}$, where f is the number of the fluoroscopic image. After that, the 2D marker positions $(\hat{c}^k)^{jif}$ were computed. Finally, to measure the 2D marker local-

[5] Discussion of such an approach can be found in the literature, e.g., [159].

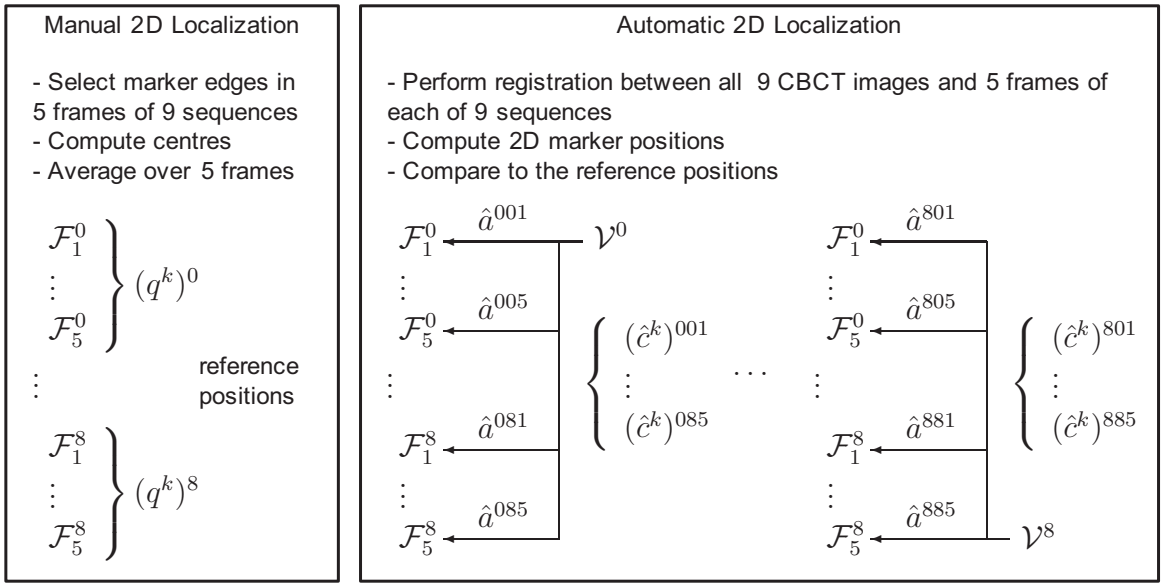


Figure 7.31: Protocol for estimation of accuracy of the 2D marker localization procedure.

ization accuracy of our automatic method in comparison to the manual localization, $(\hat{c}^k)^{jif}$ are compared to the reference positions, $(q^k)^i$, by the means of the *fiducial registration error* (FRE) [57, 58], which is defined as the root-mean-square between the reference and localized marker positions:

$$\text{FRE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \|(\hat{c}^k)^{jif} - (q^k)^i\|^2}. \tag{7.3.19}$$

Figure 7.32 presents the FRE results for different template image generation methods (the MS, VS, and MM DRR) and dimensionality of the geometric transformation (2D and 3D). For comparison, we also provide corresponding distribution for the manual localization, for which the measurements were defined as the FREs between the marker positions in each frame and the 2D reference marker positions (averages over five frames). As can be seen in Figure 7.32 and Table 7.24, all FRE values for all methods fell below the pixel size, which measures 0.2604 mm in our fluoroscopic images, and 75% of them are around or below the half of a pixel size.

The FREs for the registrations performed with the 3D geometric transformation are typically somewhat smaller than with the 2D. The exception is the MM DRR method, for which the use of the 2D and 3D transformations gives very similar results. This is due to the errors in estimation of 3D marker orientations, which, though very small (the largest error was around 3.5° for the CBCT images with 0.5 mm voxels,

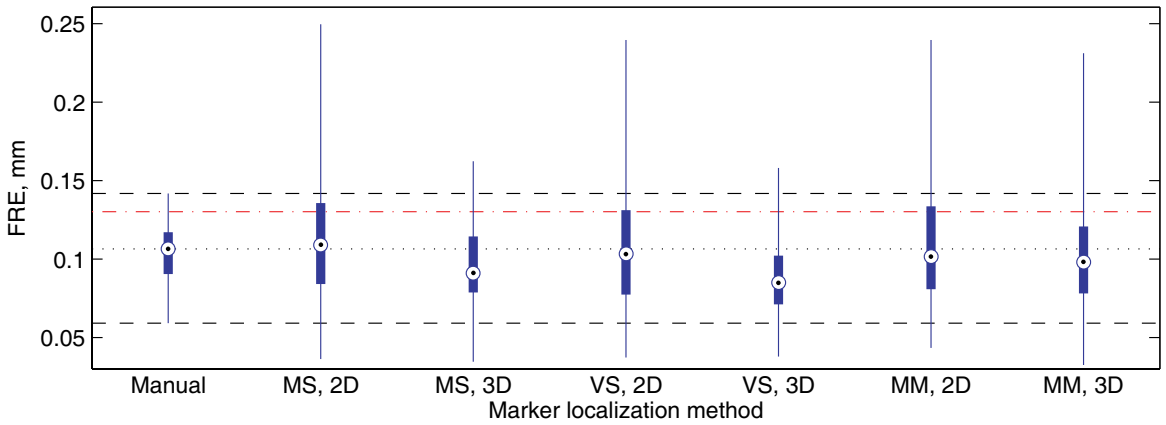


Figure 7.32: 2D marker localization accuracy for the RSVP-S data set in comparison to manual localization by the means of fiducial registration error (FRE). The *dotted line* shows the median level, and the *dashed lines* the minimal and maximal levels of the FRE distribution corresponding to the manual localization. The *dashed-dotted line* shows the level of a half of a pixel size. The corresponding numerical values are given in Table 7.24.

Table 7.24: Fiducial registration error (FRE) for manual and automatic localization results shown in Figure 7.32.

Localization	FRE, mm					
method	Min	25%	Med	75%	Max	Mean ± Std
Manual	0.06	0.09	0.11	0.12	0.14	0.1 ± 0.02
MS, 2D	0.04	0.08	0.11	0.14	0.25	0.11 ± 0.04
MS, 3D	0.03	0.08	0.09	0.11	0.16	0.1 ± 0.02
VS, 2D	0.04	0.08	0.1	0.13	0.24	0.11 ± 0.04
VS, 3D	0.04	0.07	0.08	0.1	0.16	0.09 ± 0.02
MM, 2D	0.04	0.08	0.1	0.13	0.24	0.11 ± 0.04
MM, 3D	0.03	0.08	0.1	0.12	0.23	0.1 ± 0.03

Section 7.1.7), can introduce some geometric uncertainty during registration.

A detailed analysis revealed that the largest FREs for the 2D transformations were attained in registrations that had the CBCT and the fluoroscopic images display the phantom in two different positions, for which the out-of-plane displacements were 5 and 10 mm. This is consistent with our earlier suggestion in Section 5.3.2 that the variation in the in-plane distance between the markers in the 2D images, should the fiducial markers be out of plane by 10 mm, may be around 0.23 mm, which is less than a pixel. Based on the results presented above, we conclude that, under our assumptions on the maximal possible magnitude of the marker motion, any of the proposed template image generation methods with either 2D or 3D geometric transformation provide submillimeter and subpixel accuracy and are comparable to manual localization, which makes them applicable in the clinical setting. The next section further compares the registration results produced with the proposed template image generation methods and 2D or 3D transformations in order to select the most appropriate combination for the use with our automatic 2D marker localization procedure.

7.3.3 2D versus 3D Geometric Transformations

In this section, we compare the results of the registration between the CBCT and fluoroscopic images using the 2D and 3D geometric transformations. Although the use of the 3D transformation offers slightly better localization accuracy (Section 7.3.2), the increased number of parameters in the optimization problem (six degrees of freedom for the 3D rigid transformation compared to three for the 2D) is associated with significant increase in the computational cost. For example, compare the average running times of an optimization procedure in our Matlab implementation shown in Table 7.25, which was executed on the system described in Appendix A using the MS DRR generation method with the 3D and 2D geometric transformations: Computational savings of more than 6 times can be achieved by using the 2D transformation compared to the 3D.

Table 7.25: Average running times of an optimization procedure for the MS DRRs with the 2D and 3D geometric transformations.

Localization	Running time, sec			
method	P1	P2	P3	RSVP
MS, 3D	3.43	3.29	3.89	2.83
MS, 2D	0.57	0.55	0.71	0.76

As mentioned in Section 5.3.2, we predicted that our imaging system might be not sensitive enough to adequately detect the out-of-plane marker displacement, which could make the use of the 3D transformation unjustifiable. To verify this, we designed the following extension to the experiment described in Section 7.3.2 (Figure 7.31). In that experiment, for the data set j, the registration was performed between the five frames of each of the nine fluoroscopic sequences $\mathcal{F}^0, \dots, \mathcal{F}^8$ and the CBCT image $\mathcal{V}^j$, resulting in transformation parameters $\hat{a}^{jif}$, where i is the index of the fluoroscopic sequence, and f is the number of the fluoroscopic image. Knowing these parameters for the 3D geometric transformation in the data set j, we aim to estimate the 3D marker positions in the CBCT images $\mathcal{V}^i$, $i = 0, \dots, 8$, $i \neq j$. To do so, we apply the 3D transformation with the found parameters to the 3D marker positions in the image $\mathcal{V}^j$, denoted by $(c^k_{3\mathrm{D}})^j$, that are known from the 3D marker segmentation step (Chapter 3), i.e., $\mathcal{T}^{3\mathrm{D}}\left(\hat{a}^{jif}, (c^k_{3\mathrm{D}})^j\right)$. Then, because $\hat{a}^{jif}$ includes both an estimated displacement of the phantom and a correction for the geometrical miscalibration of the system (Appendix B.4), the miscalibration has to be compensated for by applying an inverted 3D transformation with the parameters $\hat{a}^{iif}$, i.e., the transformation parameters between $\mathcal{F}^i_f$ and $\mathcal{V}^i$ coming from the data set i. Hence, the 3D marker positions in the image $\mathcal{V}^i$ can be estimated from the data set j, fluoroscopic image f, as:

$$(\hat{c}^k_{3\mathrm{D}})^{jf}_{\mathcal{V}^i} = \left(\mathcal{T}^{3\mathrm{D}}\right)^{-1}\left(\hat{a}^{iif}, \mathcal{T}^{3\mathrm{D}}(\hat{a}^{jif}, (c^k_{3\mathrm{D}})^j)\right), \tag{7.3.20}$$

where the inverse rigid transformation is defined in Appendix B.3.

After the points $(\hat{c}^k_{3\mathrm{D}})^{jf}_{\mathcal{V}^i}$ are computed, we compare them to the known 3D marker positions in the image $\mathcal{V}^i$, $(c^k_{3\mathrm{D}})^i$. The results are displayed in Figure 7.33, where the marker positions are shown separately for the X (out-of-plane displacements), Y and Z axes (in-plane displacements). We only present the results produced for the MS DRR method, as those for the VS and MM DRR are similar. The *expected positions* refer to $(c^k_{3\mathrm{D}})^i$, and the *observed* to $(\hat{c}^k_{3\mathrm{D}})^{jf}_{\mathcal{V}^i}$. As can be seen from the figure, our imaging system is not sensitive enough to correctly estimate the out-of-plane marker displacements. This observation can also be reiterated by analyzing the 3D and 2D marker trajectories consisting of the localized marker positions in the consecutive images of the fluoroscopic sequences. Figure 7.34 shows the XYZ (3D transformation) and the UV (2D transformation) marker displacements between the consecutive images along the corresponding axes for the P1, P2, P3, and RSVP data sets described in Section 7.3.1. It is easy to see that the marker displacements along the X axis are estimated to be much larger than those along the Y and Z axes. This contradicts a number of studies reporting the smallest prostate motion in the left-right patient direction (Section 5.3.1), and the fact that the phantom was stationary in this experiment. On the other hand, the distributions of the displacements along the Y

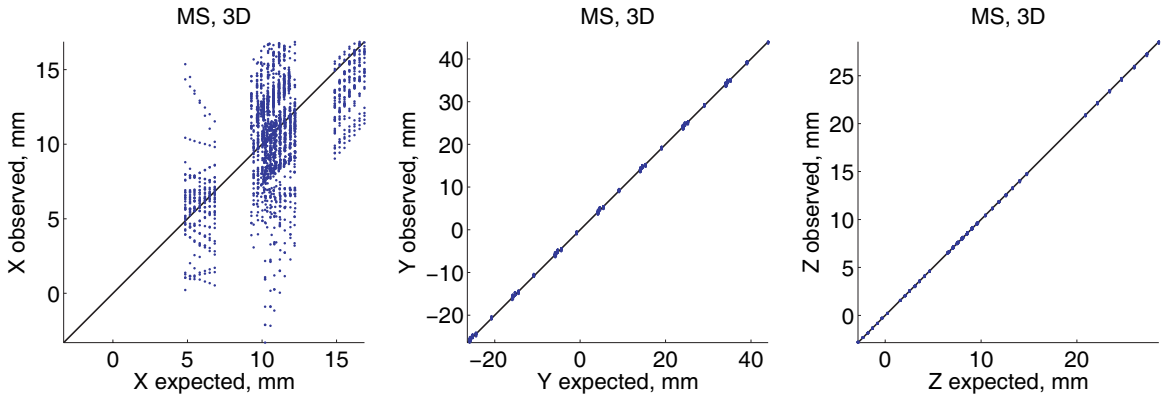


Figure 7.33: While the in-plane displacements along the Y and Z axes are accurately estimated (*centre and right*), our imaging system is not sensitive enough to detect the out-of-plane displacements along the X axis (*left*).

and Z, and U and V axes (in-plane motion) are similar, where the Y axis corresponds to the negative V, and Z to the negative U, and their magnitudes are in agreement with our observations of the marker motion in the testing data sets. Therefore, we conclude that the use of the 2D rigid geometric transformation is appropriate for the proposed marker localization procedure.

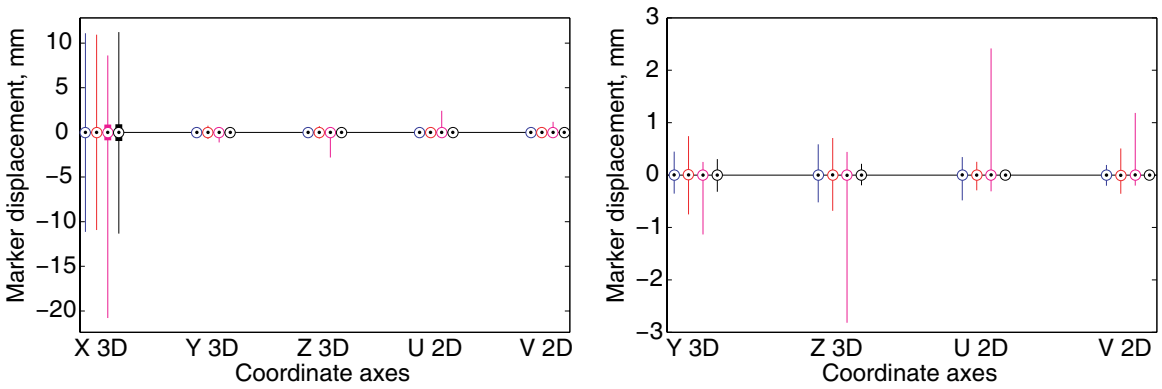


Figure 7.34: Distributions of the marker displacements along the coordinate axes between consecutive fluoroscopic images. For each axis, the four distributions correspond to the displacements in the P1, P2, P3, and RSVP data sets (*from left to right*). The inability to correctly estimate the out-of-plane motion is reflected in unreasonably large jumps along the X axis. The displacements along the Y and Z axes, corresponding to the in-plane motion, are similar to those along the U and V axes, where Y corresponds to the negative V, and Z to the negative U (*zoom-in on the right*).

7.3.4 Success Rates of Template Image Generation Methods

The average running times of the optimization procedure for the P1, P2, P3, and RSVP data sets described in Section 7.3.1 in our Matlab implementation executed on the system described in Appendix A with the use of the 2D rigid transformation are presented in Table 7.26. The MS DRR was the most efficient computationally, closely followed by the VS DRR. The use of the MM DRR is the most expensive. However, to place these numbers into the context of the whole marker localization procedure, note that the optimization is only run once for each 2D localization. It means, that given that a typical time for the whole 2D localization procedure for a patient CBCT and fluoroscopic image pair takes around 7 sec with the use of the MS DRR in our implementation, it is likely to take about 7.4 sec with the VS DRR, and 12 sec with the MM DRR. In other words, while the VS DRR may take almost twice as much time as the MS DRR, the whole localization procedure will only take a fraction of a second longer.

The localization success rates, i.e., the percentage of correct localizations, for the P1, P2, P3, and RSVP data sets for each template image generation method are given in Table 7.27. While for the P1 and RSVP the localization success rates reached 100% with all image generation methods, the lower rates for the P2 and P3 data sets require some explanation. As was discussed in Section 5.5, one of the fluoroscopic se-

Table 7.26: Average running times of an optimization procedure for the MS, VS, and MM DRRs with the 2D rigid transformation.

Localization	Running time, sec			
method	P1	P2	P3	RSVP
MS, 2D	0.57	0.55	0.71	0.76
VS, 2D	0.91	0.92	1.07	1.49
MM, 2D	5.11	5.57	6.34	6.43

Table 7.27: 2D marker localization success rates for the MS, VS, and MM DRRs with the 2D rigid transformation.

Localization	Localization success rates, %			
method	P1	P2	P3	RSVP
MS, 2D	100	100	99.39	100
VS, 2D	100	96.66	99.51	100
MM, 2D	100	99.81	99.39	100

quences in the P3 data set contained 5 frames of extremely poor quality, for which the manual marker localization by the human observer was virtually impossible. Localization failures in these images cause the overall success rate drop to 99.39% (for the VS DRR, the markers were localized correctly in one of these five frames, resulting in a slightly higher success rate of 99.51%). For all other images in the P3, the 2D marker localization was successful. As for the P2 data set, one of the CBCT images (with 0.5 mm voxels) had large streak reconstruction artifacts (false high-intensity pixels surrounding the 3D markers in the CBCT image, see Section 3.1.3). Therefore, in the VS DRR, there was a high-intensity spot next to the actual marker, the matching of which to the marker in the MEF-image produced a local minimum. As a result, the found marker positions were shifted more than the allowed 4 mm from the projected 3D marker positions (the number was selected based on the known extent of marker motion in the patient testing data sets). While the use of the MS or MM DRRs for this sequence also resulted in the shifted marker positions, those were accepted as the correct localizations, as the shift was about half the size: The centre of the generated marker (2D Gaussian) was placed in the middle between the actual marker and its artifact, which was closer to the actual measurement. The use of the MM DRR for the P2 data set resulted in an incorrect localization for only three frames out of 1619

(99.81% success rate) due to a local minimum, where one of the markers was matched correctly, but the second one registered to a thin bright anatomical feature in the vicinity of the real marker. As a result, the second and especially the third localized marker positions were too far from the estimated positions to be accepted.

To conclude, the ability to generate the markers in the template image that closely resemble those in the MEF-image, such as produced by the VS and MM DRR methods, has a potential benefit of precise localization. In particular, matching of similar intensity shapes is expected to produce smaller values of $\mathcal{D}_{\text{NNCC}}$ comparing to matching of the elongated markers in the MEF-image to the roundish shapes of the Gaussians generated by the MS DRR method. However, the localization precision can be potentially reduced in the presence of streak artifacts and errors in estimation of maker orientation in 3D (Section 7.1.7). Therefore, to process the images similar to those in our current testing data sets, we select the MS DRR method as the most reliable and efficient in our Matlab implementation, while the VS and MM DRR methods can successfully be used for the CBCT images of better quality.

7.3.5 Success Rates Depending on Optimization Starting Points

In Section 5.3.3, we discussed a number of approaches for the selection of the starting point, a^0, for the optimization method used in 2D-3D image registration. These methods are: multiresolution with the Gaussian pyramid starting with a simple projection (MR), normalized 2D cross-correlation with multiple templates (NCC-MT), NCC-MT on reduced resolutions (NCC-MT at 3×3 and 5×5 pixels), and multiresolution with the Gaussian pyramid and NCC-MT (MR + NCC-MT). To justify the choice of the method for selection of a^0, we designed the following experiment. The 2D-3D image registration was performed between the CBCT and fluoroscopic images of the P1-S, P2-S, P3-S, and RSVP-S data sets described in Section 7.3.1. The MS DRR template generation method and 2D rigid transformation were used. As we are interested in designing a method that has a sufficiently large *capture range*, we have artificially moved the markers in 3D prior to registration by 20 mm in the YZ plane, which is the largest shift allowed in-plane. The registration results are evaluated in terms of *success rates*, i.e., the percentage of correct registrations, and an *average running time* of one registration in our Matlab implementation executed on the system described in Appendix A. The registration procedure includes the definition/computation of a starting point, forming of multiresolution levels (if necessary), and the optimization procedure pertaining to this registration.

The results for P1-S, P2-S, P3-S, and RSVP-S data sets are presented in Tables 7.28 – 7.35, respectively. The success rates and running times are provided de-

pending on the number of fluoroscopic frames, ℓ, used in Dolph-Chebyshev temporal filtering prior to the MEF application. We observed the following:

- Although working well for the RSVP-S data set, the 6-level MR method never reached a 100% success rate for any of the patient testing data sets. The reason is illustrated in Figures 7.35 and 7.36, where we show the MEF- and the template images before the registration (*left*) and the corresponding images at each multiresolution level (*right, numbered*) before the registration at each level is performed. While the correct displacement for the RSVP images under registration is recovered at the coarsest level, and the starting point at the level 2 is already close to the solution (Figure 7.35), the patient images may be corrupted by more noise and anatomical measurements that survived filtering. Application of the Gaussian smoothing amplifies both markers and anatomical features, and misregistration can happen at the coarsest levels, misleading the registration at the lower levels (Figure 7.36). This, in addition to a significantly larger running time in comparison to all other methods we tested, makes this method inadequate for our marker localization procedure.
- The remaining four methods that used NCC-MT showed 100% success rates on test images for all values of ℓ. NCC-MT on full resolution and MR + NCC-MT methods have similar running times, which are consistently larger than those of the NCC-MT on reduced resolutions, for both 3×3 and 5×5 pixel averaging approaches. The use of reduced resolution results in an implementation that is 3–5 times faster.
- The time difference between the 3×3 and 5×5 pixel averaging methods is not dramatic. The 5×5 is slightly faster for most data sets, on the order of 0.1 second, except for the P3-S data set, where additional time required to form a coarse resolution image overweighs the savings resulting from executing the NCC-MT on reduced resolution. Taking into account a small marker size, 3×0.8 mm, which in our fluoroscopic images results in about 12×3 pixel measurement, we consider the NCC-MT method with 3×3 pixel averaging to be the most reasonable choice: The one that provides efficient computation but does not oversmooth the markers.

Table 7.28: Localization success rates for P1-S data set.

ℓ, frames	Success rates, %				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	48	100	100	100	100
2	66	100	100	100	100
3	46	100	100	100	100
4	74	100	100	100	100
5	72	100	100	100	100
6	70	100	100	100	100
7	82	100	100	100	100
8	82	100	100	100	100
9	82	100	100	100	100
10	86	100	100	100	100
11	84	100	100	100	100

Table 7.29: Average running times for P1-S data set.

ℓ, frames	Running time, sec				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	12.19	2.48	0.8	0.69	3.01
2	11.38	2.46	0.76	0.66	2.9
3	11.58	2.45	0.76	0.65	2.76
4	11.2	2.46	0.73	0.66	2.81
5	11.9	2.45	0.72	0.64	2.95
6	11.4	2.46	0.73	0.65	2.83
7	11.44	2.45	0.74	0.64	2.77
8	11.24	2.5	0.74	0.65	2.87
9	11.27	2.45	0.74	0.63	2.81
10	11.3	2.44	0.74	0.63	2.79
11	10.89	2.45	0.74	0.66	2.77

Table 7.30: Localization success rates for P2-S data set.

ℓ, frames	Success rates, %				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	90	100	100	100	100
2	90	100	100	100	100
3	90	100	100	100	100
4	90	100	100	100	100
5	90	100	100	100	100
6	90	100	100	100	100
7	90	100	100	100	100
8	90	100	100	100	100
9	90	100	100	100	100
10	90	100	100	100	100
11	90	100	100	100	100

Table 7.31: Average running times for P2-S data set.

ℓ, frames	Running time, sec				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	11.1	2.37	0.71	0.63	2.79
2	10.88	2.38	0.73	0.64	2.79
3	11.3	2.37	0.71	0.66	2.9
4	11.19	2.38	0.71	0.66	2.81
5	11.05	2.37	0.7	0.66	2.88
6	11.2	2.39	0.7	0.65	2.83
7	11.54	2.37	0.75	0.65	2.75
8	11.1	2.33	0.74	0.66	2.83
9	11.25	2.32	0.76	0.65	2.87
10	11.1	2.33	0.76	0.64	2.86
11	10.91	2.34	0.75	0.66	2.75

Table 7.32: Localization success rates for P3-S data set.

ℓ, frames	Success rates, %				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	82	100	100	100	100
2	86	100	100	100	100
3	84	100	100	100	100
4	84	100	100	100	100
5	88	100	100	100	100
6	90	100	100	100	100
7	86	100	100	100	100
8	90	100	100	100	100
9	96	100	100	100	100
10	98	100	100	100	100
11	98	100	100	100	100

Table 7.33: Average running times for P3-S data set.

ℓ, frames	Running time, sec				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	13.84	2.33	0.58	0.64	2.77
2	13.18	2.32	0.55	0.61	2.81
3	12.49	2.32	0.56	0.63	2.64
4	13.14	2.33	0.56	0.62	2.79
5	12.94	2.31	0.55	0.62	2.75
6	13.26	2.32	0.55	0.61	2.66
7	13.27	2.34	0.55	0.61	2.68
8	13.15	2.31	0.55	0.61	2.79
9	13.44	2.35	0.55	0.61	2.64
10	12.49	2.41	0.54	0.62	2.74
11	13.26	2.3	0.56	0.62	2.78

Table 7.34: Localization success rates for RSVP-S data set.

ℓ, frames	Success rates, %				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	100	100	100	100	100
2	100	100	100	100	100
3	100	100	100	100	100
4	100	100	100	100	100
5	100	100	100	100	100
6	100	100	100	100	100
7	100	100	100	100	100
8	100	100	100	100	100
9	100	100	100	100	100
10	100	100	100	100	100
11	100	100	100	100	100

Table 7.35: Average running times for RSVP-S data set.

ℓ, frames	Running time, sec				
	MR	NCC-MT (full)	NCC-MT (3×3)	NCC-MT (5×5)	MR + NCC-MT
1	13.82	3.19	0.98	0.85	3
2	13.25	3.21	0.92	0.81	2.83
3	13.08	3.24	0.89	0.78	2.78
4	12.8	3.24	0.88	0.78	2.73
5	13.29	3.28	0.9	0.78	2.74
6	12.62	3.23	0.89	0.78	2.51
7	12.85	3.2	0.88	0.75	2.62
8	12.32	3.26	0.89	0.75	2.69
9	12.63	3.21	0.87	0.76	2.77
10	12.46	3.21	0.88	0.77	2.68
11	12.24	3.12	0.88	0.76	2.67

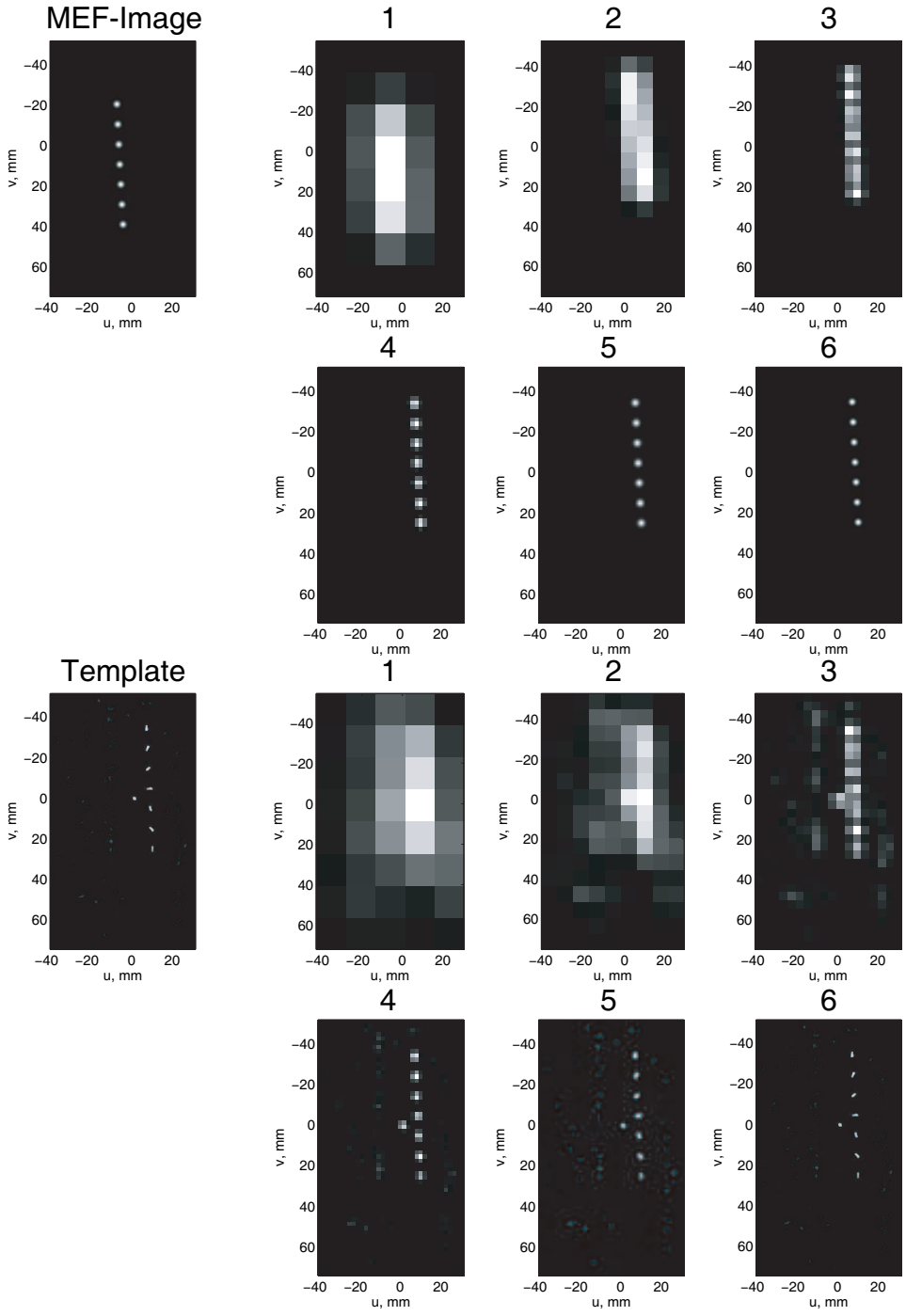


Figure 7.35: The initial MEF- and template images of the RSVP (*left*) and the corresponding images at each multiresolution level (*right, numbered*) before the registration was performed. The use of multiresolution leads to 100% registration success rates for the testing RSVP data set, where a significant initial displacement is typically recovered at the coarsest multiresolution level, and the solution is improved at the lower levels.

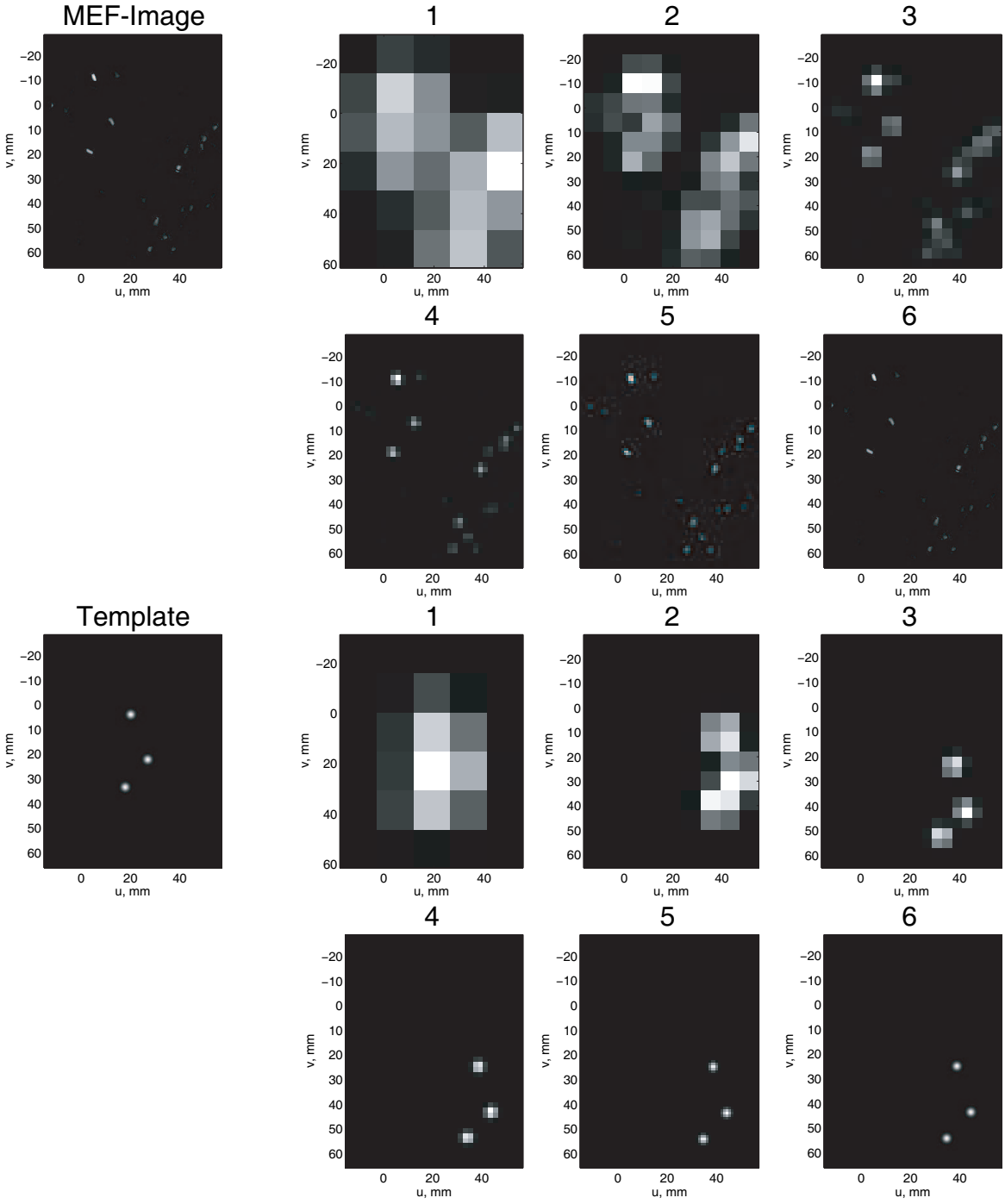


Figure 7.36: The initial MEF- and template images of a prostate patient (*left*) and the corresponding images at each multiresolution level (*right, numbered*) before the registration was performed. In some patient images, the prominent anatomical feature that survived the MEF filtering can cause registration failure at the highest levels, thereby misleading registration at the lower levels.

7.3.6 Computational Cost Depending on Temporal Filtering

As explained in Section 5.4, the combination of the methods in the proposed marker localization procedure and experiments in Sections 7.3.2–7.3.5 led to 100% registration success rates on our testing data sets. Therefore, the registration success itself can not be a deciding factor in the selection of the filter width, ℓ, of the temporal Dolph-Chebyshev filter that will be used before the application of the MEF.

We executed the complete 2D marker localization procedure on all images of the P1-S, P2-S, P3-S, and RSVP-S data sets described in Section 7.3.1, consisting of the total of 50 images for the patient and 45 for the phantom data sets. The total 2D localization running times of our Matlab implementation on the system described in Appendix A for different values of ℓ, divided by the total number of images in each data set, are presented in Table 7.36. The reduction in time for the increasing values of ℓ is due to diminishing computational time of the MEF processing and optimization: It typically takes less time to filter and register smoother images.

We select $\ell = 5$ as being small enough to not significantly reduce temporal resolution of the markers while large enough to improve the CNR. Filtering five frames acquired at 5.5 Hz produces an image that contains measurements gather over an interval of just less than one second.

Table 7.36: Average running times for the 2D localization procedure per image.

ℓ, frames	Running time, sec			
	P1-S	P2-S	P3-S	RSVP-S
1	21.16	7.23	12.42	153.77
2	13.33	5.74	7.57	60.93
3	11.53	5.44	6.90	40.46
4	8.34	5.31	6.26	28.53
5	9.40	5.32	6.11	24.23
6	8.40	5.29	5.94	21.27
7	8.52	5.56	5.75	18.91
8	8.36	5.54	5.59	17.13
9	8.20	5.64	5.55	15.66
10	8.18	5.60	5.58	14.56
11	7.76	5.62	5.68	12.24

7.4 Marker Motion Tracking

7.4.1 Testing Data Set for Tracking Validation

In this section, we describe the data sets used to select an appropriate registration approach in Section 7.4.2 and to select, train, and validate a prediction model in Sections 7.4.3 – 7.4.6. The models under consideration are described in Section 6.2 and include zeroth-order prediction, linear extrapolation, and linear regression.

In total, we have acquired 23 fluoroscopic image sequences where the phantom was performing different types of motion in the direction along the V axis. The traces are shown in Figures 7.37, 7.38, and 7.39, and can briefly be described as follows:

- The phantom was moving back and forth between two positions along the V axis. It gradually slowed down when approaching extreme points, and accelerated towards the middle. The trace of the motion in time was sinusoidal, with the peak-to-peak amplitudes of 2, 4, 10, and 20 mm, and the periods of 2.5, 5, 10, 15 sec, resulting in 16 sequences (Figures 7.37 and 7.38).
- The phantom was programmed to move according to three breathing traces of patients selected from the library supplied with the phantom control software: fast, slow, and an irregular motion traces (Figure 7.39, *left column*).
- The phantom remained stationary in one of the sequences.
- The phantom performed constant speed linear motion between two positions with slow, medium, and fast speed (Figure 7.39, *right column*).

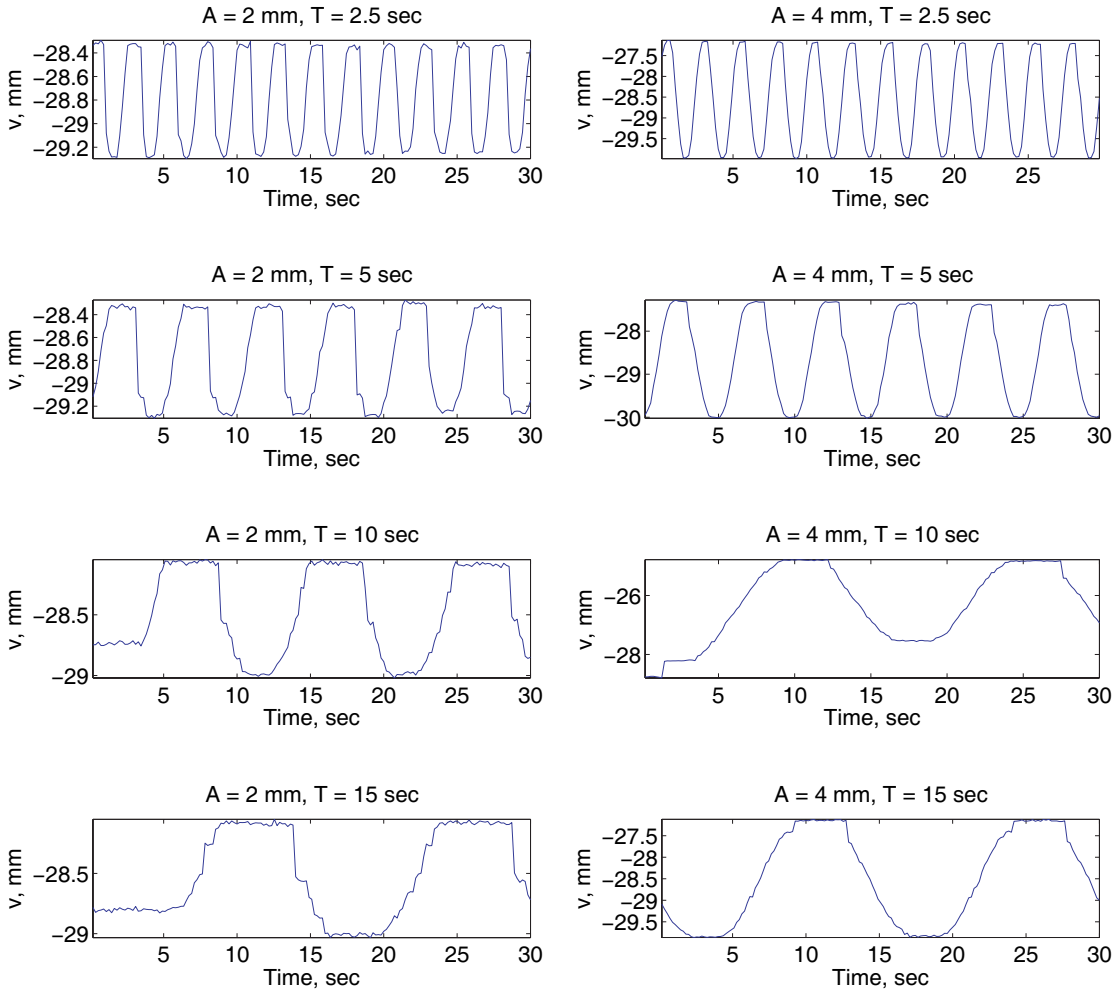


Figure 7.37: Motion traces of the Quasar phantom: sinusoidal motion with the peak-to-peak amplitudes of 2 and 4 mm and periods of 2.5, 5, 10, and 15 sec.

All sequences had 164–166 images, except for one sequence of 92 images, resulting in 3,723 images in total. The traces in Figures 7.37, 7.38, and 7.39 are produced by following one of the seven markers. As can be seen from the figures, the non-ideality in the phantom physical build, including the use of the custom insert with the fiducial markers instead of the standard inserts supplied with the phantom, resulted in non-ideality of the traces. Therefore, the tracking results had to be compared to the positions of the markers directly extracted from the 2D images, rather than to the programmed phantom motion traces.

Some of the experiments were also conducted using patient testing data sets. They included ten fluoroscopic sequences for Patient 1, ten for Patient 2, and five for

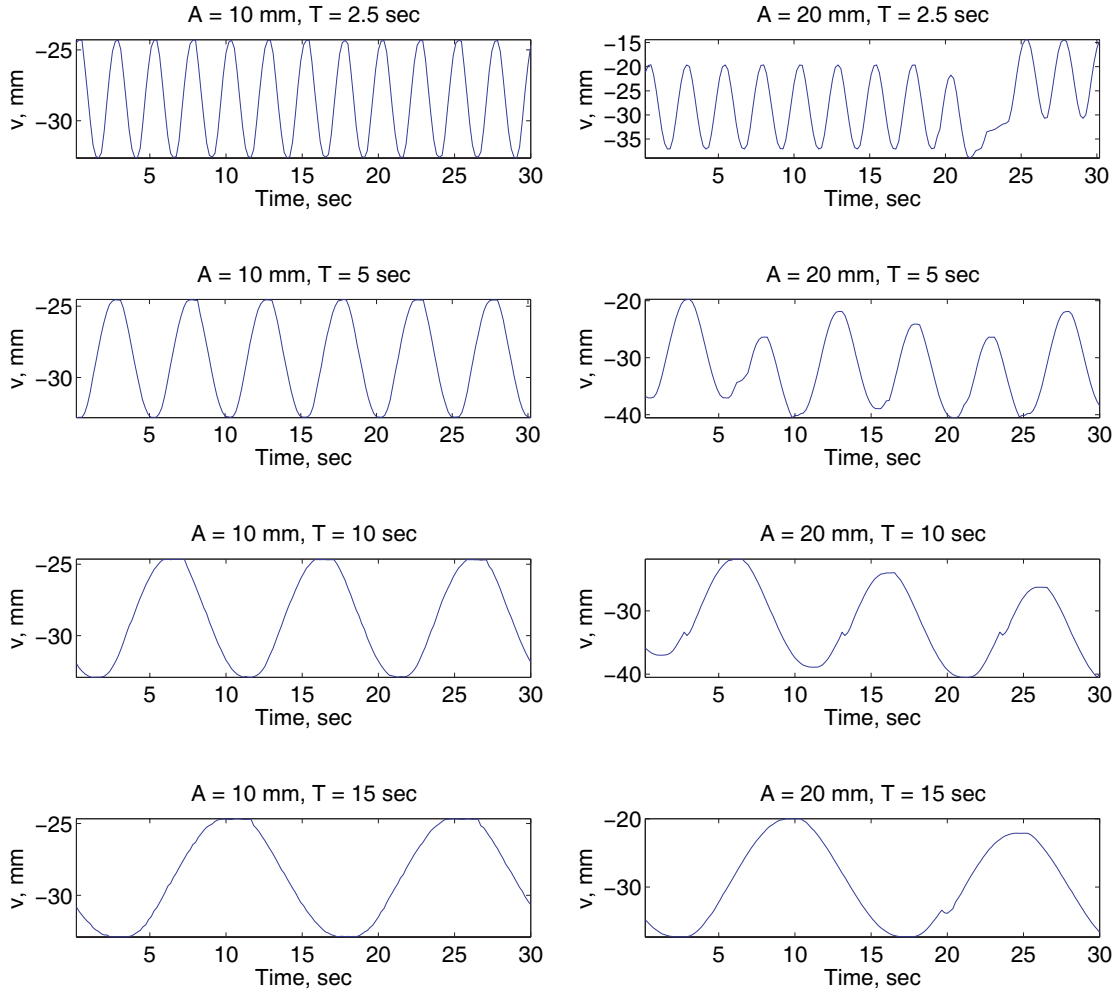


Figure 7.38: Motion traces of the Quasar phantom: sinusoidal motion with the peak-to-peak amplitudes of 10 and 20 mm and periods of 2.5, 5, 10, and 15 sec.

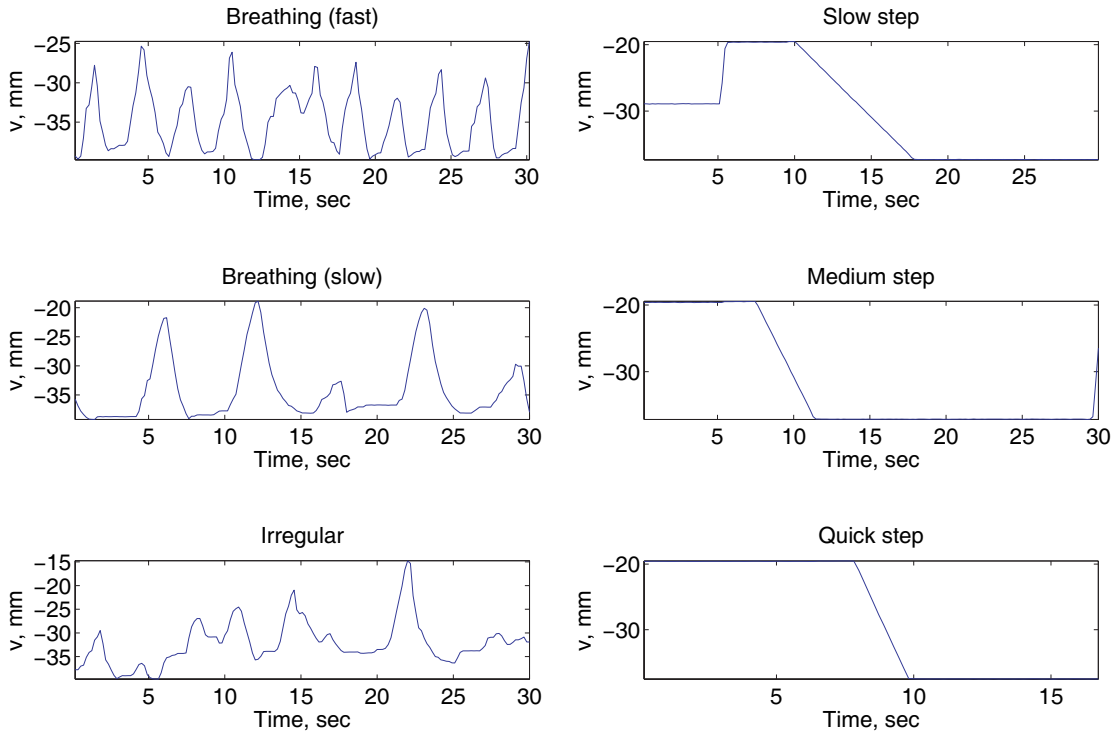


Figure 7.39: Motion traces of the Quasar phantom: fast, slow, and irregular breathing of a patient (*left column*), and slow, medium, and fast linear motion (*right column*).

Patient 3, around 160 images each. In Section 7.3, these data sets were referred to as P1, P2, and P3, respectively.

7.4.2 Registration Approach for Tracking Procedure

In order to choose an appropriate image registration approach for the tracking procedure, we have performed marker tracking in several fluoroscopic image sequences of the Quasar motion phantom from the data set described in Section 7.4.1. The two approaches introduced in Section 6.1 that are compared in terms of computational cost and tracking accuracy are (i) the direct registration of the images to the template (first image of the sequence) by using different interpolation schemes and (ii) the preprocessing of the images in the sequence prior to registration using the marker enhancement filter (MEF, Section 4.4) or its modification, the magnitude-only filter (MOF, Section 4.4), and generating the template with the 2D variation of the marker splat (MS) method discussed in Section 6.1. Temporal filtering has not been used as the markers in the fluoroscopic images of the Quasar phantom possessed sufficient contrast with the background.

We first discuss the approach of image registration with the use of interpolation, where no preprocessing was performed on the images. The first image of the sequence was used as the template. The parameters of the geometric transformation for the markers found in the preceding image were used as the prediction for the current image (zeroth-order predictor, Section 6.2), and defined the tracking ROIs as described in Sections 6.1 and 4.3. As described in Section 6.3, in order to decrease the probability of an optimization routine becoming stuck in a local minimum in the process of registration, the prediction was further improved by the shift-only normalized cross-correlation method. The application of the 2D rigid geometric transformation to the template was performed by applying commonly used interpolation schemes, such as the nearest neighbour, linear, cubic, and cubic spline [150, 69].

Table 7.37 shows running times of tracking in our Matlab implementation on the system described in Appendix A, applied to three image sequences of the Quasar motion phantom, each of which has 164 images. The sequences 1, 2, and 3 correspond to the medium step (one-time constant speed linear motion between two points), sinusoidal with 10 mm amplitude and 10 seconds period, and irregular patient breathing sequences described in Section 7.4.1. In addition to the four interpolation schemes, mentioned above, we also provide timing of the 2D marker localization performed independently in each image of the sequence in the *last column*. As can be seen from the presented results, while the nearest neighbour method incurs a lower computational cost than the independent 2D localization, the other methods incur similar or substantially larger costs. Due to a large range of the running times for the different

Table 7.37: Tracking running times for three fluoroscopic image sequences of the Quasar motion phantom. The registration is performed using interpolation, without prior marker enhancement, and compared to the independent 2D marker localization in each image of the sequence.

	Running times, hh : mm : ss				
	Interpolation scheme				
Sequence number	Nearest neighbour	Linear	Cubic	Cubic spline	Independent 2D localization
1	00:18:39	03:08:01	05:14:43	02:27:53	00:20:54
2	00:05:41	00:16:12	01:07:52	01:04:59	00:20:23
3	00:06:56	00:21:51	00:27:12	01:10:42	00:20:53

methods, in this section we report results in the hours, minutes, and seconds format (hh:mm:ss), rather than the usual seconds format.

In the search for a computationally efficient approach, we turned to another option, namely using the 2D marker splat (MS) method to generate the template image (Section 6.1) coupled with preprocessing of the fluoroscopic images by the MEF (Section 4.4.1). The difference between such a model and the independent 2D localization is that the marker positions from the preceding image are used to form a tracking ROI to perform image registration in the subsequent image, which is smaller than the localization ROI. Also, in the process of registration, the 2D rigid transformation is applied directly to the 2D marker positions in the template, without the use of the 3D image. As can be seen from Table 7.38 (2^{nd} *column*), the use of such a scheme presents the possibility for moderate computational savings in comparison to the independent 2D localization (Table 7.37, *last column*).

As explained in Section 6.1, the ability to predict the 2D marker positions and to use small tracking ROIs (in comparison to the larger localization ROIs) can provide opportunity for further improvement in the efficiency by using a reduced version of the MEF, namely the MOF (Section 4.4.5). Similar to tracking with the full MEF, the template in the process of registration is generated by the 2D MS method. The results presented in Table 7.38 (*last column*) demonstrate a significant improvement in computational efficiency.

Note that the use of the MS and MEF/MOF methods (both in tracking and localization) results in similar execution times for different fluoroscopic sequences (with the same number of images, and the same parameters used for ROI computation). This is due to the creation of the images that possess very similar intensities and fea-

Table 7.38: Tracking running times for three fluoroscopic image sequences of the Quasar motion phantom. The registration is performed with the application of the MEF or MOF preprocessing and the 2D MS template image generation method.

Sequence number	Running times, hh : mm : ss	
	MEF	MOF
1	00:16:04	00:03:32
2	00:15:58	00:03:37
3	00:16:56	00:03:36

Table 7.39: The medians of the fiducial registration error (FRE) in comparison to the independent 2D marker localization for three fluoroscopic sequences of the Quasar motion phantom.

Sequence number	Fiducial registration error, mm					
	Interpolation scheme					
	Nearest neighbour	Linear	Cubic	Cubic spline	MEF	MOF
1	0.0294	0.1572	0.1552	0.1552	0.0002	0.0003
2	0.0200	0.1529	0.1591	0.1530	0.0002	0.0003
3	0.0279	0.1444	0.1460	0.1460	0.0003	0.0004

tures, unlike in the registration approaches that use interpolation, where the initial fluoroscopic images can differ by noise levels and marker contrast, thus resulting in different times required for optimization.

Finally, we have also analyzed marker tracking accuracy of all methods described above in comparison to the marker positions obtained with the independent 2D localization. Figure 7.40 shows the distributions of the fiducial registration error (FRE, Equation (7.3.19)) [57, 58], which we define here as the root-mean-square between the localized and tracked marker positions. While all methods demonstrate subpixel accuracy for most of the images (most FRE measurements are below the *dashed line*), the methods using digital preprocessing with MEF/MOF and 2D MS template generation showed the best agreement with the independent localization. The FRE medians for the three Quasar sequences are given in Table 7.39.

Based on computational efficiency of our Matlab implementation, and sup-

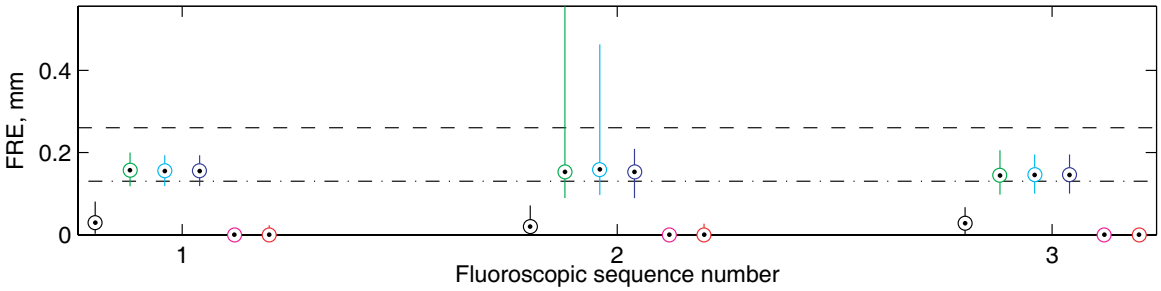


Figure 7.40: Tracking accuracy in terms of fiducial registration error (FRE) for three fluoroscopic sequences of the Quasar motion phantom, compared to the independent 2D marker localization. For each sequence, the six distributions (*from left to right*) correspond to the FRE results where the registration was performed by using the nearest neighbour, linear, cubic, spline interpolation, the MEF/MS, and the MOF/MS. The *dashed line* shows the level of a pixel size, and the *dashed-and-dotted line* shows the level of the half pixel size. The corresponding FRE median values are given in Table 7.39.

ported by the tracking accuracy results, we use the 2D MS template generation method coupled with the MOF preprocessing of the images in the 2D-2D image registration procedure of our tracking module.

7.4.3 Linear Regression Training

In an ideal situation, if the system can be exactly described by a linear regression (LR) model, there should be coefficients $\beta = (\beta_0, \ldots, \beta_n)^T$ such that the transformation parameters for the i^{th} image, $(a^*)^i$, can be expressed as $\beta_0 + \beta_1 (a^*)^{i-1} + \ldots + \beta_n (a^*)^{i-n}$. Since the geometric parameters (shifts and rotations) that describe marker motion are not expected to be precisely defined by a linear system, and we would like to be able to determine one set of the parameters β to be used for all images, the coefficients β are computed as an approximate solution of an overdetermined system of equations, where the number of variables in the system equals $n+1$, n is the number of preceding images used in prediction, and the number of equations is the total number of images, for which the training is done, multiplied by three, as there is one equation for each geometric transformation parameter, $a = (t_u, t_v, \gamma)^T$. For 23 sequences described in Section 7.4.1, the number of equations was more than 10^4. The solution β can be found by minimizing the sum of squared differences between the predicted parameters $\hat{a}^i_\beta = \beta_0 + \beta_1 (a^*)^{i-1} + \ldots + \beta_n (a^*)^{i-n}$ (Equation (6.2.7)) and the optimized parameters

Table 7.40: Coefficients β for the linear regression model obtained from training on 23 image sequences of the Quasar motion phantom. The number of images (*first column*) refers to the number of the preceding images that are used to predict the transformation parameters in the current image.

Number of images	Linear regression coefficients							
	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
2	0	1.8	−0.83					
3	0	1.99	−1.23	0.23				
4	0	1.95	−1.07	−0.02	0.12			
5	0	1.92	−1.08	0.29	−0.43	0.28		
6	0	1.87	−1	0.24	−0.2	−0.12	0.21	
7	0	1.87	−1	0.23	−0.19	−0.12	0.19	0.01

$(a^*)^i$ for all i in the following problem:

$$\min_{\beta} \sum_i \|(a^*)^i - \hat{a}^i_\beta\|^2. \tag{7.4.21}$$

The optimized parameters a^* of the geometric transformation were found by the independent 2D marker localization in each image of each sequence as described in Chapter 5.

Table 7.40 presents the values of β obtained by solving (7.4.21). For any number of preceding images that we tested, the coefficient β_0 for the constant term was very small (on the order of 10^{-4}), and the largest contribution to the prediction was obtained from the geometric parameters of the two closest in time images, $(a^*)^{i-1}$ and $(a^*)^{i-2}$. Note that for the LR models with different number of preceding images, the largest coefficients are similar. In fact, the coefficients are close in values to the those of the linear extrapolation (LE) model defined in Section 6.2 ($\beta_0 = 0$, $\beta_1 = 2$, and $\beta_2 = -1$, see Equation (6.2.6)).

7.4.4 Validation of Prediction Models on Phantom Data

In this section, we compare the prediction models defined in Section 6.2 in application to Quasar motion phantom data set described in Section 7.4.1. The models that we evaluated were: zeroth-order prediction (ZO), linear extrapolation (LE), and linear regression (LR) model with different numbers of preceding images (see Section 7.4.3).

To compare the prediction models, we define a *fiducial prediction error* (FPE) similar to the fiducial registration error (FRE, Equation (7.3.19)) [57, 58]. For each

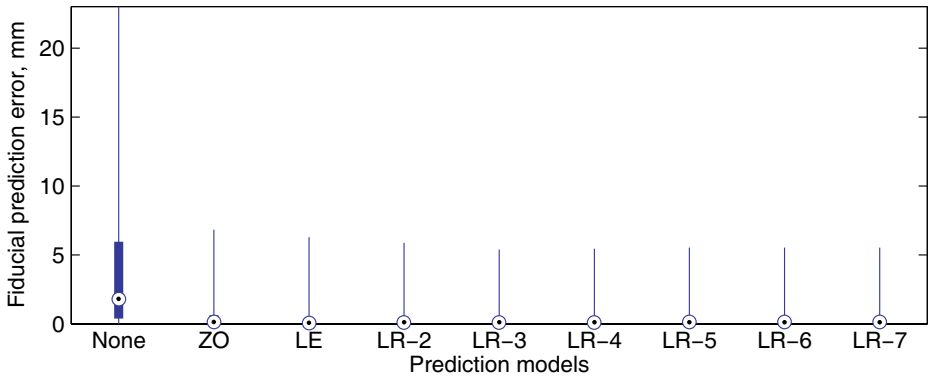


Figure 7.41: Distributions of the fiducial prediction error (FPE) for phantom images for different prediction models. Table 7.41 provides the explanation for the abbreviations used and FPE numerical values.

image of each sequence, let $q^k, k = 1, \ldots, n$ be the known positions of n markers, obtained by the independent 2D marker localization in every image, and $\hat{c}^k$ be the marker positions predicted by one of the models. Then, the FPE of the given model is computed as:

$$\text{FPE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n} \|\hat{c}^k - q^k\|^2}. \tag{7.4.22}$$

The FPE results for the rest of the models in application to the 23 Quasar image sequences (3,723 images in total) described in Section 7.4.1 are presented in Figure 7.41 and Table 7.41. The "No prediction (None)" distribution corresponds to the overall marker motion between the template created from the first image of the sequence and all subsequent images of that sequence. As can be seen in Table 7.41, the ZO and LE models are characterized with very similar results, and improve most of the statistical indicators 10-fold and maximal FPE around 3.5 times in comparison to no prediction. The LR model coefficients β used in this experiment were obtained as described in Section 7.4.3 by a "leave-one-out" approach: For each sequence, the vector β is computed from the 22 sequences, excluding the sequence of interest. We experimented with different number of preceding images used in the LR model, from 2 to 7. The FPE results are very similar, which is explained by the fact that the corresponding coefficients β_i are similar regardless of the number of preceding images used, and the two largest coefficients are assigned to the geometric parameters of the two preceding images (see more details in Section 7.4.3). The median FPE of the LR approach is similar to that of the ZO and LE models (0.04 mm difference corresponds to around 1/6 of a pixel size), however, the maximal FPE is improved by around

Table 7.41: Distributions of the fiducial prediction error (FPE) for phantom images for different prediction models.

Prediction model abbreviation	Fiducial prediction error, mm					Prediction model description
	Min	25%	Med	75%	Max	
None	0	0.39	1.81	5.96	23.01	No prediction
ZO	0	0.03	0.15	0.64	6.83	Zeroth order prediction
LE	0	0.04	0.07	0.24	6.29	Linear extrapolation
LR-2	0	0.05	0.11	0.31	5.87	Linear regression, 2 images
LR-3	0	0.05	0.11	0.32	5.39	Linear regression, 3 images
LR-4	0	0.05	0.11	0.32	5.46	Linear regression, 4 images
LR-5	0	0.06	0.12	0.29	5.54	Linear regression, 5 images
LR-6	0	0.05	0.13	0.31	5.54	Linear regression, 6 images
LR-7	0.01	0.05	0.13	0.32	5.54	Linear regression, 7 images

1–1.5 mm (4–6 pixels).

The results reported above showed that all considered models reduced the FPE significantly comparing to no prediction. The LR models were better than ZO and LE in reducing the maximal FPE. Therefore, we select the LR model based on three preceding images (LR-3) as a predictor in our marker tracking module, as it provides slightly better results than the rest of the LR models. The coefficient vector used in the proposed tracking system was $\beta = (0, 1.99, -1.23, 0.23)^T$ (see Section 7.4.3). Our choice of the prediction model was also validated by the experiments with the patient testing data set, the results of which are reported in Section 7.4.5.

7.4.5 Validation of Prediction Models on Patient Data

In addition to testing different prediction models on image sequences of the Quasar motion phantom, as described in Section 7.4.4, we have also applied the models to the patient image sequences described in Section 7.4.1 to verify the findings. The images were preprocessed by the temporal Dolph-Chebyshev filter with the kernel size of $\ell = 5$. In addition, we computed a patient-based LR model, denoted by PB in the graphs. In this model, coefficients β were obtained from training the model on the patient data with the "leave-one-out" approach, i.e., for each patient, the coefficients β were found by training the model on the images of the two remaining patients. The distributions of the FPE are shown in Figure 7.42 and corresponding median and maximal FPE values are given in Table 7.42. The outliers in Figure 7.42 (*bottom*)

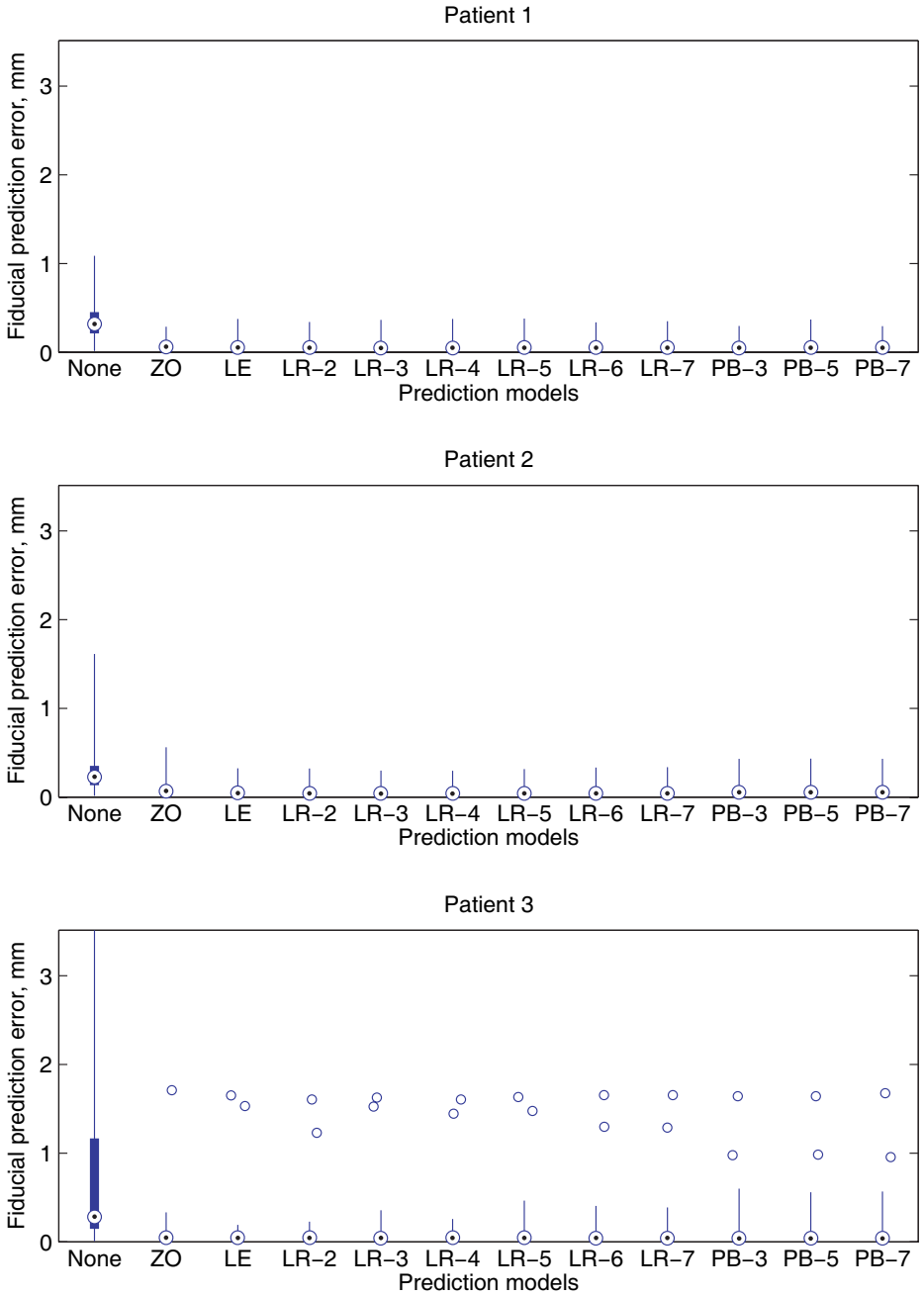


Figure 7.42: Distributions of the FPE for patient images for different prediction models. Table 7.42 provides the corresponding FPE numerical values.

Table 7.42: Distributions of the FPE for patient images for different prediction models.

	Fiducial prediction error, mm						
	Patient 1		Patient 2		Patient 3		
Prediction model	Med	Max	Med	Max	Med	Max	Maximal outlier
None	0.32	1.09	0.23	1.61	0.28	3.51	3.51
ZO	0.06	0.29	0.07	0.56	0.05	0.33	1.71
LE	0.05	0.37	0.05	0.32	0.05	0.19	1.65
LR-2	0.05	0.34	0.04	0.32	0.05	0.23	1.61
LR-3	0.05	0.36	0.04	0.3	0.04	0.36	1.63
LR-4	0.05	0.38	0.04	0.3	0.05	0.26	1.6
LR-5	0.05	0.38	0.04	0.32	0.05	0.47	1.63
LR-6	0.05	0.34	0.04	0.33	0.04	0.4	1.65
LR-7	0.05	0.35	0.04	0.34	0.04	0.39	1.66
PB-3	0.05	0.3	0.05	0.43	0.04	0.6	1.64
PB-5	0.05	0.37	0.05	0.43	0.04	0.56	1.64
PB-7	0.05	0.29	0.05	0.43	0.04	0.57	1.64

correspond to the images where the markers performed rapid moves according to the following scenario: The markers remained nearly stationary up until and including image i, then moved by approximately 2 mm between the images i and $i+1$ (the prediction for image $i+1$ was close to the optimal transformation parameters of image i, the first outlier), and after that remained almost stationary (but the prediction moved further along the difference between the images i and $i+1$, the second outlier). The analysis of the FPE distributions leads to conclusions similar to the ones presented in Section 7.4.4: All models improve the FPE comparing to no prediction. However, the difference between the models is not very significant. Based on the experiment described above, the choice of LR-3 is confirmed to be reasonable. The patient-based LR models produced similar (in many cases slightly higher) errors than the LR-3 model: The maximal FPE for different patients and number of images used in prediction ranged from 0.29 to 0.6 mm compared to 0.3 to 0.36 mm for the LR-3 model.

7.4.6 Accuracy of Marker Motion Tracking

The validation of the tracking method accuracy was performed by computing the fiducial registration error (FRE, Equation (7.3.19)) between the marker positions found by the tracking procedure and by the independent 2D localization performed in every im-

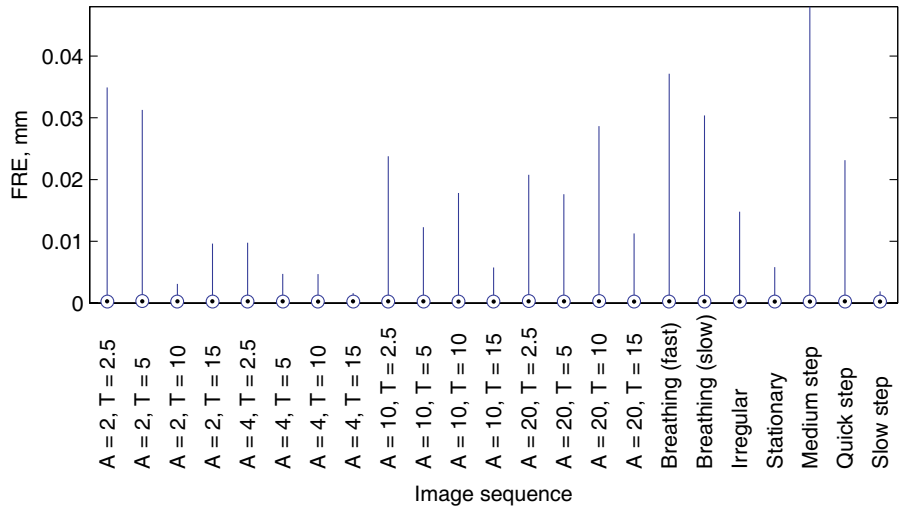


Figure 7.43: Accuracy of marker tracking measured by the distributions of the FRE in fluoroscopic image sequences of the Quasar moving phantom.

age of the sequence. The accuracy of the marker localization method was evaluated in Section 7.3.2, and showed a median FRE of 0.11 mm and a maximal FRE of 0.25 mm (smaller than a pixel size) on the testing data set, where manual localization was used as the "ground truth". Figure 7.43 presents the FRE distributions for 23 fluoroscopic image sequences of the Quasar motion phantom described in Section 7.4.1. The FRE results for the Quasar phantom demonstrate high agreement between the marker positions found by the tracking and the independent 2D localization procedures. For all sequences, the minimal FREs were on the order of 10^{-5} mm, the medians were on the order of 10^{-4} mm, and the maxima were ranging from $1.6 \cdot 10^{-3}$ to $4.8 \cdot 10^{-2}$ mm. For all sequences combined, the median FRE was $2.7 \cdot 10^{-4}$, and the mean $\pm$ one standard deviation was $(5.62 \pm 24) \cdot 10^{-4}$ (Table 7.43). The FRE magnitude did not seem to have any particular relation to the trajectory and amplitude of the motion in our experiment, although it tended to be slightly higher for higher speeds (shorter period). Due to the submillimeter and subpixel accuracy, the proposed tracking method can be implemented for clinical use.

The validation of the tracking procedure on the patient testing data set consisting of 25 fluoroscopic image sequences described in Section 7.4.1 shows similar results: Although the FRE magnitudes presented in Figure 7.44 and Table 7.43 are larger in comparison to the FREs in the phantom image sequences, all maximal FREs are within a pixel size, and median or mean FREs are much smaller than a millimeter.

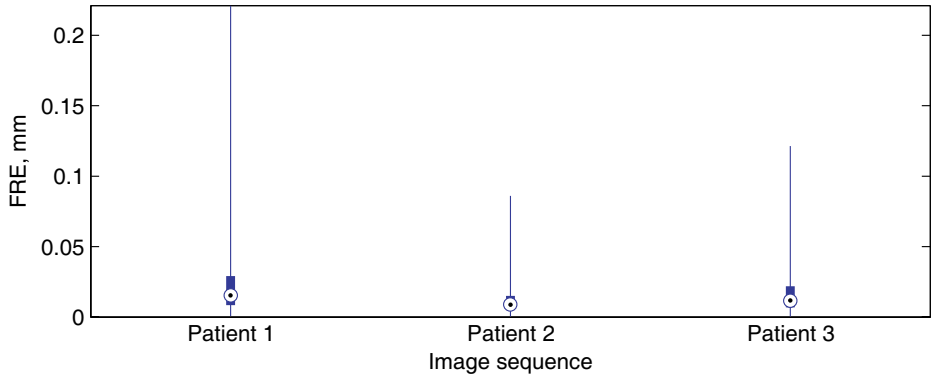


Figure 7.44: Accuracy of marker tracking measured by the distributions of the FRE in patient fluoroscopic image sequences. The corresponding numerical values are given in Table 7.43.

Table 7.43: Distributions of the FRE in Quasar phantom and patient image sequences.

	Fiducial registration error, mm					
Image sequence	Min	25%	Med	75%	Max	Mean $\pm$ Std
Quasar	0	0	0	0	0.05	0 ± 0
Patient 1	0	0.01	0.02	0.03	0.22	0.02 ± 0.03
Patient 2	0	0.01	0.01	0.02	0.09	0.01 ± 0.01
Patient 3	0	0.01	0.01	0.02	0.12	0.02 ± 0.02

Chapter 8

Patient Studies

This chapter presents the results obtained by applying the marker localization and tracking methods proposed in this thesis and implemented in the Matlab package that we called Gryphon. Examples are provided of both inter-fraction marker displacement statistics extracted from the daily position verification CBCTs in Section 8.1, and intra-fraction marker motion results computed from the fluoroscopic image sequences of the prostate patients in Section 8.2.

8.1 Inter-Fraction Marker Displacement from CBCT Images

After describing the CBCT patient data sets in Section 8.1.1, we demonstrate how the 3D positions of the markers localized in the CBCT images of a patient or a patient cohort are used to analyze inter-fraction anatomy changes in Section 8.1.2. In Section 8.1.3, we give an example of organ motion statistics that could be used to customize the radiation treatment.

8.1.1 Patient CBCT Data Set

For inter-fraction displacement analysis, we have used CBCT images of three prostate patients. Each patient had two CBCT images acquired per treatment fraction, as was briefly mentioned in Section 2.2. First, the image we denote as CBCT_s was obtained after the patient had been set up on the LINAC couch with the help of the room lasers and immobilization devices. It was used to measure the displacement of the target relative to the treatment plan, and to perform necessary corrections by moving the couch. Then, another image, CBCT_e, was acquired after the treatment procedure was completed. If the target had moved too far from its initial position, it could mean

that the prescribed radiation dose had not been delivered, and some adjustments to the treatment regimen could be necessary.

The patients in our data set came in for the treatment every working day (Monday through Friday) during the period of around two months, which resulted in 39 fractions for Patient 1, and 36 fractions for both Patient 2 and 3. For every fraction, we have both $CBCT_s$ and $CBCT_e$ images. All patients had three gold cylindrical fiducial markers, 3×0.8 mm in size, inserted into their prostates.

Ideally, to measure an inter-fraction motion, the positions of the markers in the daily CBCT images should be compared to their positions in the planning CT image. However, we did not have this information available. Instead, we used the positions of the markers in the $CBCT_e$ image of the first treatment fraction, which we expect to be the closest to the ones in the planning CT among all other CBCT images of each patient data set. First, this fraction is the closest in time to planning, implying the hope for small anatomical changes. Second, the markers in the $CBCT_s$ image from the same fraction are subject to the displacement between the CT and $CBCT_s$ (inter-fraction motion), while $CBCT_e$ image is acquired after the couch correction has been performed and may include some organ displacements that have taken place during the treatment fraction (intra-fraction[1] motion), which, according to multiple studies summarized by Byrne [38], tend to be smaller than the inter-fraction motion. Finally, the goal of this chapter is to demonstrate the application of the system to real patient data. We aim to provide examples on the kind of analysis that can be performed on patient images, rather than reporting clinically significant results.

8.1.2 Inter-Fraction Anatomical Changes and Marker Migration

A number of studies reported the use of *inter-marker distances* (IMD) to examine marker migration and anatomy changes during the course of radiation therapy [36, 114, 176, 190]. Similarly, our proposed 3D marker segmentation approach (Chapter 3) implemented in the Gryphon software can be used to conduct such analysis.

Figure 8.1 and Table 8.1 present the results on the IMD variations for the data sets described in Section 8.1.1. The distributions in Figure 8.1 show the change in the IMD compared to the IMDs in the $CBCT_e$ image of the first treatment fraction. The labels such as "M1/2-s" refer to the IMD variations between the markers 1 and 2 in the $CBCT_s$ images, and "All-e" to the variations of all IMDs combined in the $CBCT_e$

[1] Although most authors refer to any displacements measured within one fraction as the intra-fraction motion, sometimes a term peri-fraction is used to separate the displacement measured by two CBCT images acquired before and after the treatment fraction from a series of images acquired during the fraction.

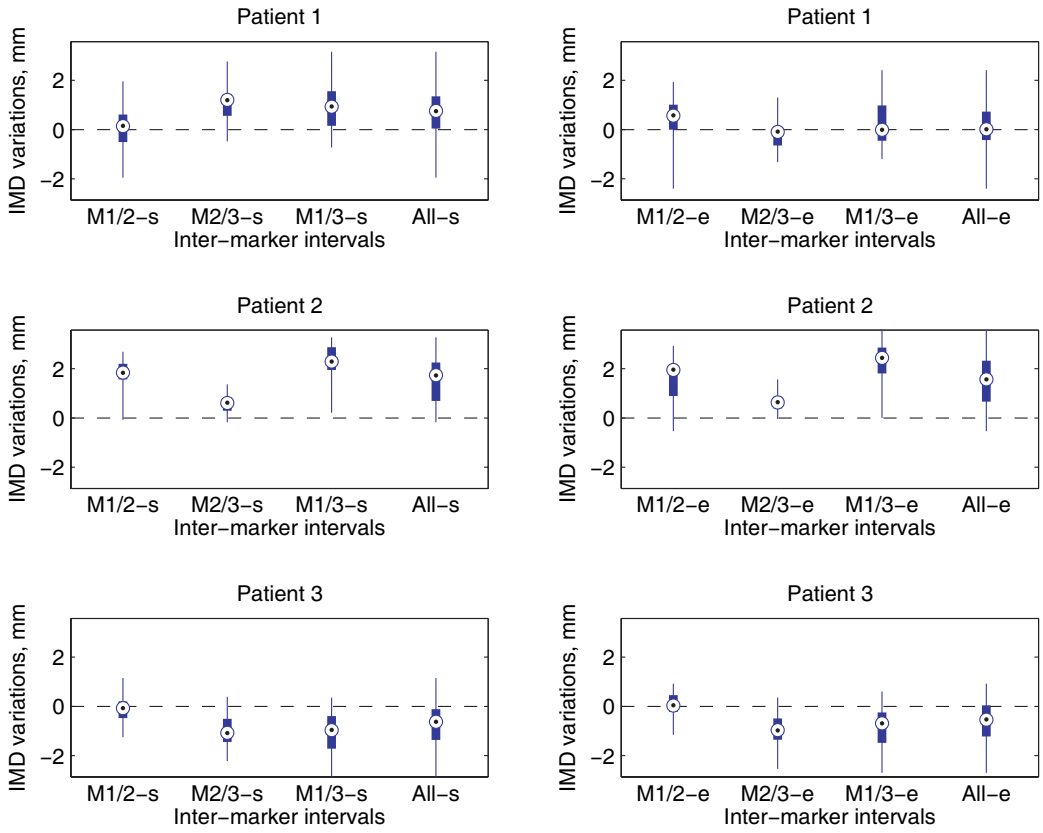


Figure 8.1: Variations in inter-marker distances (IMD) during the course of treatment.

images. Markers 1, 2, and 3 correspond to the *base*, *posterior*, and *apex* markers in the prostate patients, respectively (see Figure 7.23). We can also present the IMD results as a cumulative distribution of the absolute IMD values of all IMDs and all patients combined as shown in Figure 8.2.

Table 8.1 lists the numerical results for each patient for all IMDs combined in the pre- or post-treatment CBCT images, CBCT_s and CBCT_e. The IMD variation for all IMDs in all images of all patients combined is 0.4 ± 1.28 mm. Although based on only three prostate patients, these results are similar to several other studies. Kupelian *et al.* [114] collected statistics on the IMDs computed from a pair of x-ray images for 56 patients, 36 fractions each, and found that the average IMD variation was -0.31 ± 1.41 mm, with the largest observed change of 10.2 mm. They found that the IMD values larger than 1, 2, and 3 mm constituted 38%, 11%, and 4% of all IMDs (compare to 43%, 17%, and 3% presented in Figure 8.2, respectively). They concluded that the IMDs varied minimally, meaning little deformation of the prostate and the absence of significant marker migration. Budiharto *et al.* [36] showed that

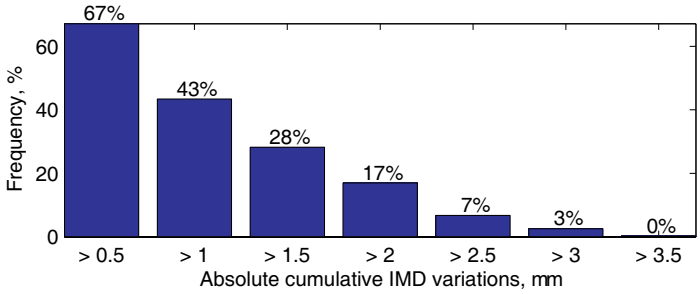


Figure 8.2: Absolute cumulative IMD variations for all patients.

Table 8.1: Variations in inter-marker distances (IMD) during the course of treatment.

	Variations in inter-marker distances, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
Patient 1, All-s	−1.95	0.04	0.75	1.35	3.16	0.7 ± 0.97
Patient 1, All-e	−2.39	−0.42	0.01	0.73	2.41	0.15 ± 0.82
Patient 2, All-s	−0.17	0.70	1.73	2.25	3.27	1.54 ± 0.93
Patient 2, All-e	−0.54	0.66	1.56	2.33	3.57	1.52 ± 1.01
Patient 3, All-s	−2.86	−1.36	−0.62	−0.11	1.15	-0.75 ± 0.81
Patient 3, All-e	−2.70	−1.22	−0.53	0.04	0.92	-0.61 ± 0.89

the number of times the variation in absolute IMDs exceeded values of 1, 2, 4, 6, and 10 mm was 33%, 14%, 5%, 3%, and 2%, respectively. The data was obtained from the orthogonal portal images of 8 patients, totalling in 219 image pairs. Poggi *et al.* [176] analyzed orthogonal radiographs of 9 patients, and found that the marker migration was 1.2 ± 0.2 mm. Schallenkamp *et al.* [190] computed IMDs of 20 patients from their portal images acquired during 39 to 41 treatment fractions, and found that they exceeded 1 and 1.5 mm in 21% and 4% of the cases.

8.1.3 Inter-Fraction Marker Displacement

The knowledge of setup errors is important in radiation therapy, in particular, the inter-fraction motion statistics can be used for safety margin design [219]. Figures 8.3 – 8.5 and corresponding numerical values of the distributions in Tables 8.2 – 8.4 represent the inter-fraction displacement results produced by using the proposed 3D marker segmentation module for the data set described in Section 8.1.1. "M1" to "M3" denote markers 1 to 3, that correspond to the *base*, *posterior*, and *apex* markers in the prostate patients, respectively (see Figure 7.23), "s" and "e" mark the data from $CBCT_s$ and $CBCT_e$ images, respectively, and "CoM" denotes a 3D centre of mass of all three markers, which we treat as the position of the target volume. Section 5.3.1 provides a brief description of the patient coordinate system, where the directions are typically called the left-right (LR), superior-inferior (SI), and anterior-posterior (AP), and correspond to the X, Y, and Z axes of the LINAC, respectively.

The displacements are computed relative to the marker positions in the $CBCT_e$ image of the first treatment fraction. As practically all displacements are smaller in the $CBCT_e$ images compared to the $CBCT_s$ (see also the cumulative inter-fraction displacement distributions for all patients in Figure 8.6), we conclude that the intra-fraction motion in the prostate patients in our data set is smaller than the inter-fraction displacements, which agrees with multiple studies (see the summary provided by Byrne [38]). The inter-fraction displacements of the centre of mass of all markers in the $CBCT_s$ images combined for all patients were -2.04 ± 2.82, 0.2 ± 3.56, 1.83 ± 3.61, and 6 ± 2.24 mm for the LR/X, SI/Y, and AP/Z directions, and in the 3D, respectively. For $CBCT_e$ images, the corresponding statistics were -0.7 ± 1.4, 0.33 ± 2.23, 1.33 ± 2.34, and 3.37 ± 1.83 mm. The maximal displacements among all patients were 4.76, 8.19, 12.06, and 14.91 mm for the LR/X, SI/Y, and AP/Z directions, and in the 3D in the $CBCT_s$ images, respectively, and 2.96, 6.12, 5.7, and 7.63 mm in the $CBCT_e$ images. For comparison, the maximal magnitudes of the inter-fraction displacements measured in ten studies reviewed in [116] cover the ranges of 2–9.3 mm for the LR/X, 4–9.9 mm for the SI/Y, 7–14 mm for the AP/Z, and 8–20 mm for the 3D displacement. Figure 8.7 shows the 3D "trajectories" of the inter-fraction marker motion in the $CBCT_s$ and

$CBCT_e$ images of the three prostate patients.

There was one outlier in the Patient 2 data set, which we removed from the reported statistics and figures. In the pre-treatment image the patient was, indeed, significantly shifted from the treatment position: The target position (centre of mass of the markers) was displaced by -1.55, -25.36, and -1.56 mm in the X, Y, and Z directions, respectively, which may indicate at error in the initial patient setup.

As pointed out by van Herk [219], the mean and standard deviation values of the inter-fraction displacement can be used to determine the setup errors, and to compute the treatment margin. While there are multiple "recipes" for computing the margins, many of them use systematic and random components of the setup error that are computed as explained below [219]. The inter-fraction displacements are measured for a number of patients undergoing multiple treatment fractions, and the mean and standard deviation (SD) values are computed for each patient across multiple fractions, such as listed in Table 8.5. Then, the group systematic error, M, is a mean of the mean values of all patients. It often deviates from zero significantly due to inaccuracy of the equipment. The SD of the systematic error, Σ, is the SD computed over all patient means. It describes the reproducibility of the treatment setup. Finally, individual patient SDs give the SD of the random error for each patient. The group mean of SDs of the random error, σ, is computed as the root-mean-square of the SDs of all patients. The results for our patient data set described in Section 8.1.1 are presented in Table 8.5. Different combinations of M, Σ, and σ have been used to estimate the radiation treatment margin [219].

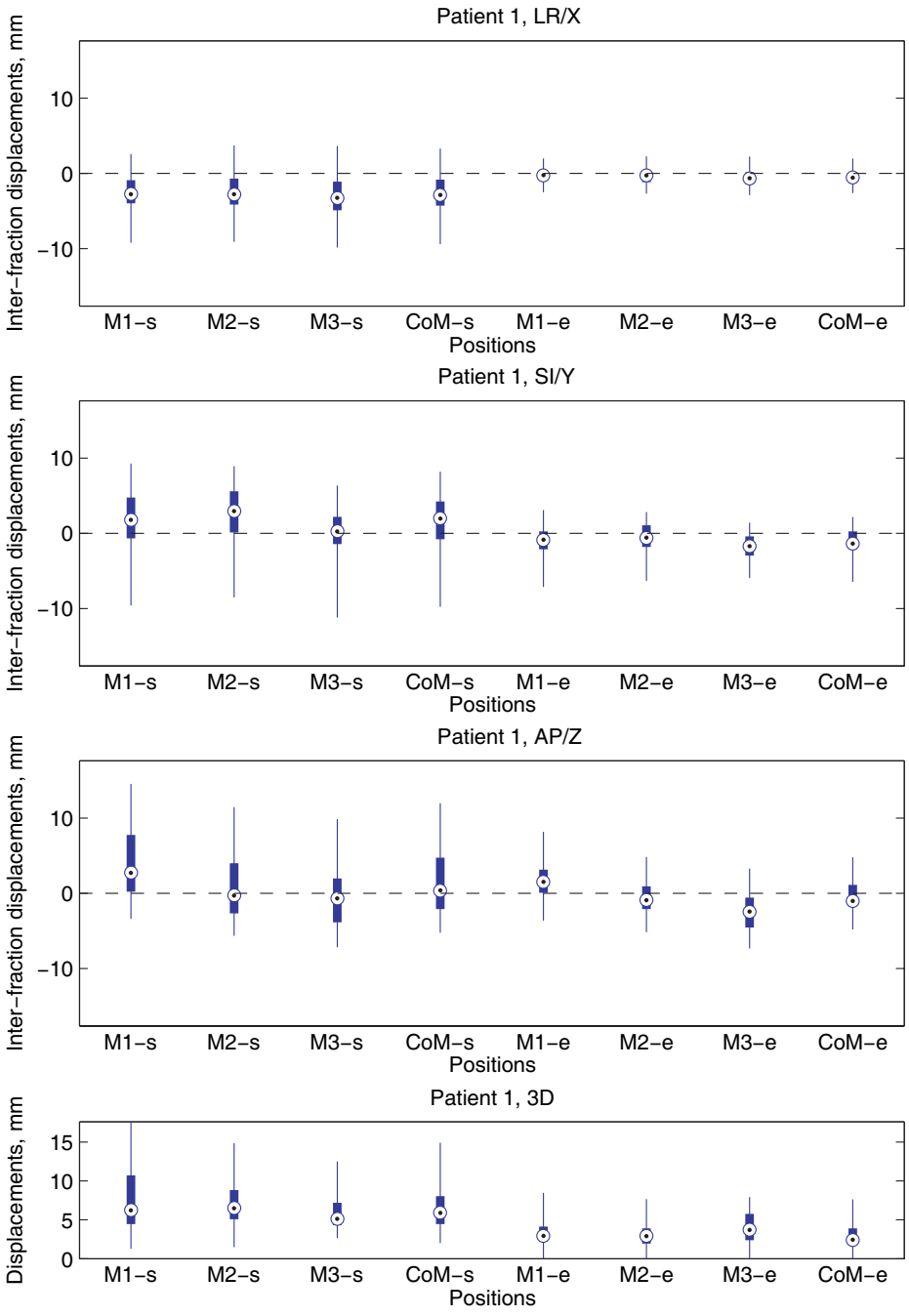


Figure 8.3: Inter-fraction displacements for Patient 1.

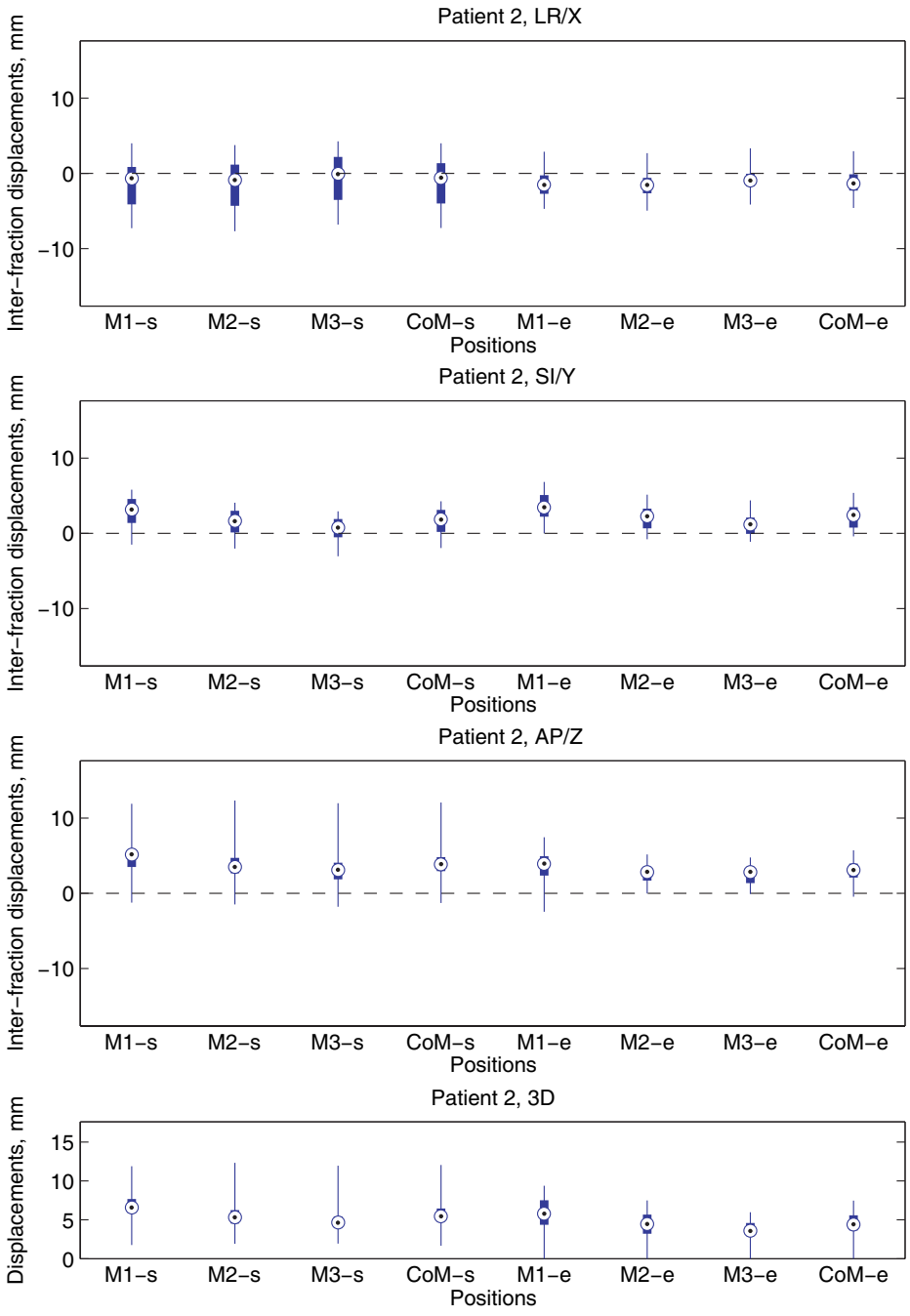


Figure 8.4: Inter-fraction displacements for Patient 2.

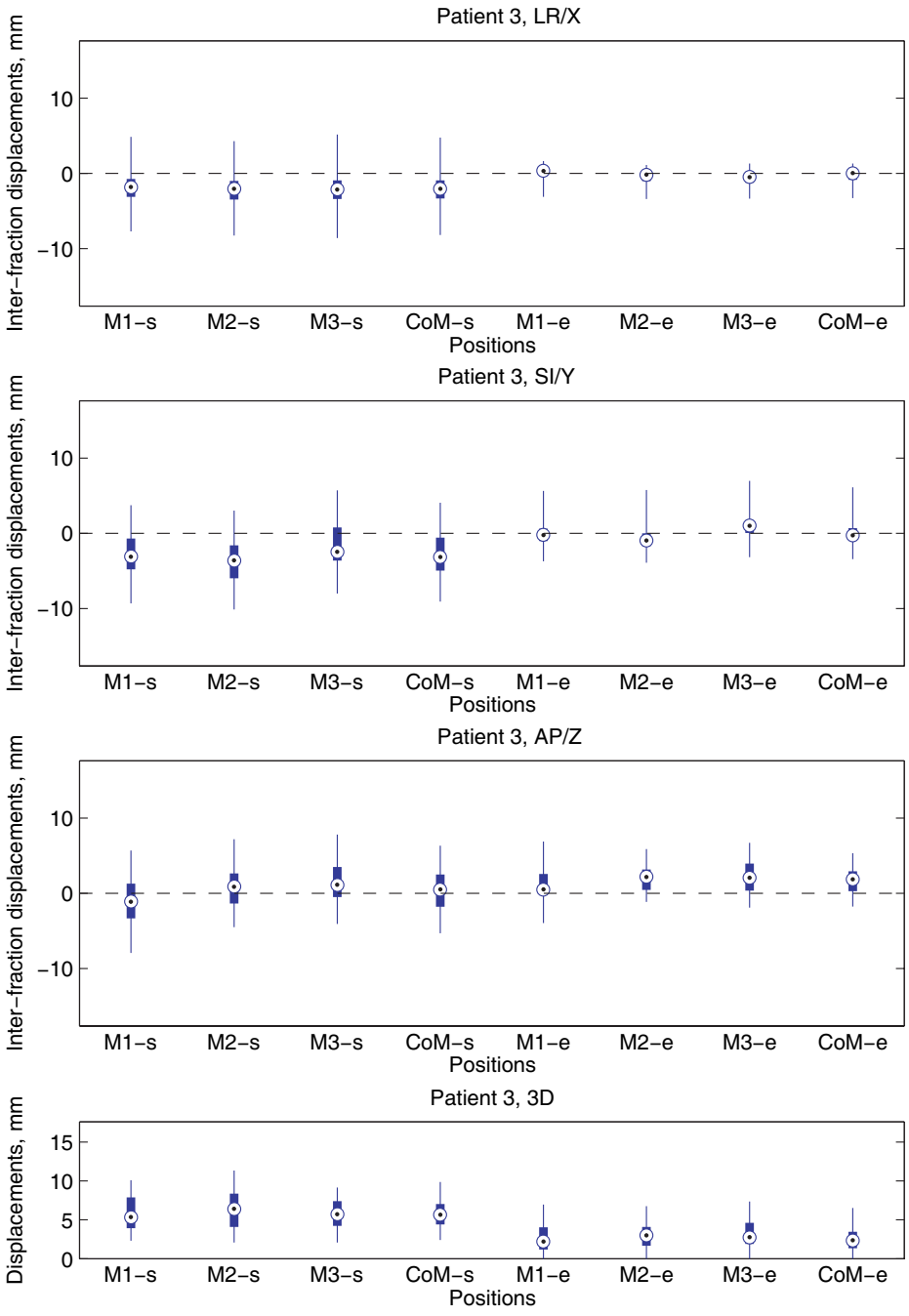


Figure 8.5: Inter-fraction displacements for Patient 3.

Table 8.2: Inter-fraction displacements for Patient 1.

	Inter-fraction displacements, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
			Patient 1, LR/X			
M1-s	−9.22	−3.97	−2.75	−0.9	2.56	−2.63 ± 2.37
M2-s	−9.1	−4.15	−2.78	−0.7	3.72	−2.59 ± 2.52
M3-s	−9.84	−4.9	−3.27	−1.09	3.64	−3.13 ± 2.62
CoM-s	−9.38	−4.29	−2.87	−0.83	3.31	−2.78 ± 2.49
M1-e	−2.49	−0.77	−0.23	0.35	1.97	−0.25 ± 0.98
M2-e	−2.7	−1.16	−0.26	0.19	2.26	−0.4 ± 1.13
M3-e	−2.89	−1.38	−0.65	0.16	2.25	−0.66 ± 1.16
CoM-e	−2.62	−1.07	−0.58	0.14	1.98	−0.44 ± 1.05
			Patient 1, SI/Y			
M1-s	−9.59	−0.67	1.78	4.76	9.28	1.79 ± 3.85
M2-s	−8.51	0.11	2.95	5.58	8.92	2.59 ± 3.68
M3-s	−11.17	−1.42	0.26	2.19	6.36	0.12 ± 3.12
CoM-s	−9.75	−0.78	1.96	4.22	8.19	1.5 ± 3.52
M1-e	−7.11	−2.12	−0.86	0.26	3.09	−0.91 ± 2.11
M2-e	−6.33	−1.8	−0.58	1.07	2.82	−0.68 ± 2.16
M3-e	−5.94	−2.95	−1.7	−0.41	1.42	−1.82 ± 1.8
CoM-e	−6.46	−2.09	−1.37	0.26	2.17	−1.13 ± 1.95
			Patient 1, AP/Z			
M1-s	−3.4	0.22	2.74	7.76	14.53	4 ± 4.97
M2-s	−5.63	−2.67	−0.28	4.01	11.45	0.92 ± 4.48
M3-s	−7.16	−3.86	−0.66	1.97	9.84	−0.47 ± 4.19
CoM-s	−5.23	−2.09	0.39	4.73	11.94	1.48 ± 4.42
M1-e	−3.65	0.08	1.52	3.13	8.16	1.83 ± 2.61
M2-e	−5.15	−2.1	−0.9	0.92	4.82	−0.77 ± 2.28
M3-e	−7.31	−4.54	−2.44	−0.58	3.28	−2.51 ± 2.52
CoM-e	−4.81	−1.67	−1.01	1.13	4.78	−0.48 ± 2.2
			Patient 1, 3D			
M1-s	1.27	4.45	6.23	10.71	17.61	7.45 ± 3.87
M2-s	1.5	5.07	6.5	8.83	14.85	6.78 ± 2.72
M3-s	2.67	4.32	5.12	7.18	12.48	6.13 ± 2.41
CoM-s	2.01	4.46	5.9	8.04	14.91	6.48 ± 2.74
M1-e	0	2.22	2.93	4.13	8.46	3.51 ± 1.99
M2-e	0	1.95	2.91	3.9	7.65	3.1 ± 1.59
M3-e	0	2.37	3.72	5.77	7.93	3.99 ± 2.21
CoM-e	0	2.08	2.42	3.91	7.63	2.99 ± 1.52

Table 8.3: Inter-fraction displacements for Patient 2.

	Inter-fraction displacements, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
			Patient 2, LR/X			
M1-s	−7.27	−4.11	−0.64	0.86	3.99	−1.26 ± 3.05
M2-s	−7.67	−4.3	−0.88	1.18	3.76	−1.31 ± 3.18
M3-s	−6.8	−3.53	−0.07	2.18	4.26	−0.43 ± 3.23
CoM-s	−7.25	−3.99	−0.6	1.36	4	−1 ± 3.15
M1-e	−4.7	−2.71	−1.51	−0.26	2.87	−1.48 ± 1.81
M2-e	−4.95	−2.65	−1.54	−0.59	2.68	−1.63 ± 1.84
M3-e	−4.14	−1.68	−0.95	−0.05	3.33	−0.87 ± 1.85
CoM-e	−4.6	−2.28	−1.33	−0.12	2.96	−1.33 ± 1.82
			Patient 2, SI/Y			
M1-s	−1.51	1.38	3.15	4.56	5.81	3.03 ± 1.89
M2-s	−2.03	0.13	1.65	3	4.06	1.6 ± 1.61
M3-s	−3.05	−0.52	0.76	1.89	2.92	0.62 ± 1.52
CoM-s	−1.96	0.2	1.83	3.13	4.27	1.75 ± 1.65
M1-e	0	2.21	3.45	5.1	6.84	3.57 ± 1.82
M2-e	−0.78	0.67	2.26	3.29	5.13	2.17 ± 1.59
M3-e	−1.12	−0.06	1.2	2.1	4.37	1.17 ± 1.46
CoM-e	−0.39	0.77	2.44	3.48	5.37	2.3 ± 1.57
			Patient 2, AP/Z			
M1-s	−1.23	3.5	5.15	5.62	11.89	4.47 ± 2.65
M2-s	−1.48	2.67	3.5	4.7	12.32	3.4 ± 2.4
M3-s	−1.77	1.85	3.11	4.07	11.97	2.99 ± 2.3
CoM-s	−1.28	2.94	3.88	4.81	12.06	3.62 ± 2.39
M1-e	−2.45	2.35	3.95	4.93	7.45	3.52 ± 2.05
M2-e	−0.01	1.69	2.82	3.47	5.16	2.66 ± 1.32
M3-e	−0.09	1.37	2.83	3.23	4.75	2.43 ± 1.24
CoM-e	−0.45	2.09	3.08	3.95	5.7	2.87 ± 1.43
			Patient 2, 3D			
M1-s	1.76	5.91	6.59	7.68	11.89	6.86 ± 1.76
M2-s	1.92	4.53	5.32	6.26	12.32	5.52 ± 1.84
M3-s	1.94	3.94	4.64	5.44	11.97	4.87 ± 1.84
CoM-s	1.66	4.85	5.48	6.45	12.06	5.66 ± 1.76
M1-e	0	4.36	5.78	7.52	9.39	5.74 ± 2.24
M2-e	0	3.21	4.45	5.67	7.49	4.37 ± 1.69
M3-e	0	2.99	3.58	4.6	5.98	3.64 ± 1.34
CoM-e	0	3.65	4.41	5.58	7.48	4.47 ± 1.74

Table 8.4: Inter-fraction displacements for Patient 3.

	Inter-fraction displacements, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
			Patient 3, LR/X			
M1-s	−7.71	−3.12	−1.8	−0.74	4.84	−1.97 ± 2.51
M2-s	−8.25	−3.47	−2.05	−1	4.29	−2.43 ± 2.61
M3-s	−8.57	−3.41	−2.14	−0.93	5.16	−2.32 ± 2.71
CoM-s	−8.18	−3.32	−2.03	−0.92	4.76	−2.24 ± 2.6
M1-e	−3.1	−0.2	0.34	0.65	1.64	0.02 ± 1.1
M2-e	−3.41	−1.03	−0.17	0.14	1.12	−0.47 ± 1.08
M3-e	−3.35	−1.23	−0.49	0.05	1.31	−0.58 ± 1.09
CoM-e	−3.29	−0.86	0.02	0.24	1.31	−0.34 ± 1.06
			Patient 3, SI/Y			
M1-s	−9.31	−4.78	−3.1	−0.7	3.74	−2.78 ± 3.08
M2-s	−10.1	−5.97	−3.58	−1.61	3.02	−3.67 ± 3.32
M3-s	−7.99	−3.61	−2.45	0.79	5.71	−1.64 ± 3.4
CoM-s	−9.05	−4.95	−3.14	−0.56	4.07	−2.7 ± 3.24
M1-e	−3.7	−1.02	−0.21	0.61	5.63	−0.17 ± 1.56
M2-e	−3.89	−1.68	−0.94	0	5.76	−0.82 ± 1.72
M3-e	−3.16	0.04	1.03	1.65	6.97	0.85 ± 1.7
CoM-e	−3.42	−0.55	−0.27	0.69	6.12	−0.05 ± 1.58
			Patient 3, AP/Z			
M1-s	−7.92	−3.33	−1.11	1.29	5.68	−0.89 ± 3.3
M2-s	−4.48	−1.35	0.89	2.63	7.19	0.82 ± 2.86
M3-s	−4.08	−0.5	1.12	3.49	7.79	1.47 ± 3.05
CoM-s	−5.31	−1.79	0.51	2.49	6.32	0.47 ± 2.93
M1-e	−3.95	−0.07	0.53	2.56	6.88	1.09 ± 2.4
M2-e	−1.15	0.45	2.19	3.17	5.88	2.02 ± 1.87
M3-e	−1.89	0.37	2.06	3.95	6.71	2.13 ± 2.14
CoM-e	−1.75	0.28	1.85	2.93	5.32	1.75 ± 1.89
			Patient 3, 3D			
M1-s	2.31	3.92	5.36	7.9	10.09	5.8 ± 2.19
M2-s	2.09	4.09	6.41	8.37	11.34	6.31 ± 2.38
M3-s	2.07	4.22	5.72	7.4	9.14	5.82 ± 1.98
CoM-s	2.41	4.4	5.65	7.03	9.86	5.8 ± 2
M1-e	0	1.2	2.2	4.05	6.97	2.65 ± 1.85
M2-e	0	1.67	3.01	4.08	6.75	3.11 ± 1.66
M3-e	0	1.91	2.75	4.61	7.36	3.25 ± 1.89
CoM-e	0	1.35	2.37	3.47	6.54	2.68 ± 1.76

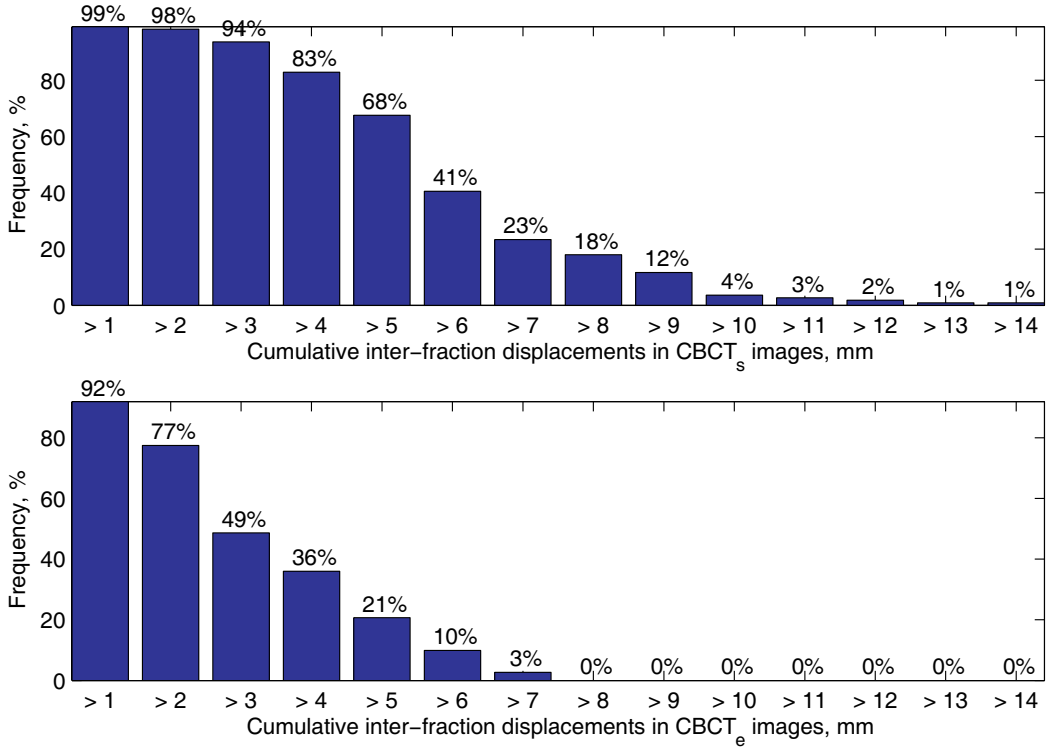


Figure 8.6: Cumulative inter-fraction motion statistics for patient CBCT images.

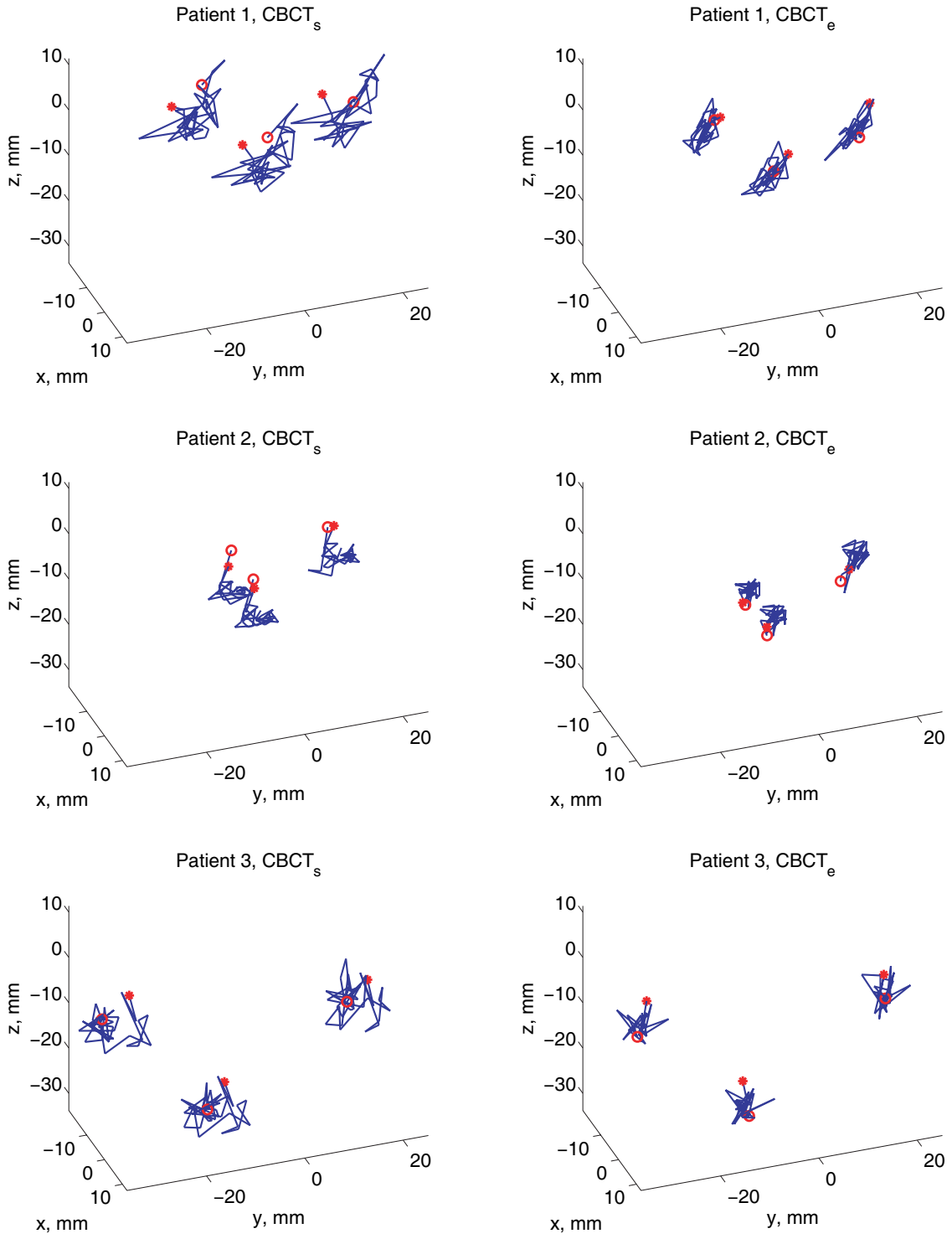


Figure 8.7: Three-dimensional trajectories of the inter-fraction motion in $\mathrm{CBCT_s}$ and $\mathrm{CBCT_e}$ images of the prostate patients. The marker positions in the first and last treatment fractions are denoted by the *circle* and *star*, respectively.

Table 8.5: Inter-fraction error statistics.

	Inter-fraction error statistics, mm					
	$CBCT_s$			$CBCT_e$		
Position	Patient 1	Patient 2	Patient 3	Patient 1	Patient 2	Patient 3
LR/X						
CoM mean	−2.78	−1.00	−2.24	−0.44	−1.33	−0.34
CoM std	2.49	3.15	2.6	1.05	1.82	1.06
	$M = -2.01$, $\Sigma = 0.91$, $\sigma = 2.76$			$M = -0.7$, $\Sigma = 0.55$, $\sigma = 1.36$		
SI/Y						
CoM mean	1.5	1.75	−2.7	−1.13	2.3	−0.05
CoM std	3.52	1.65	3.24	1.95	1.57	1.58
	$M = 0.18$, $\Sigma = 2.5$, $\sigma = 2.92$			$M = 0.37$, $\Sigma = 1.75$, $\sigma = 1.71$		
AP/Z						
CoM mean	1.48	3.62	0.47	−0.48	2.87	1.75
CoM std	4.42	2.39	2.93	2.2	1.43	1.89
	$M = 1.86$, $\Sigma = 1.61$, $\sigma = 3.36$			$M = 1.38$, $\Sigma = 1.71$, $\sigma = 1.87$		
3D						
CoM mean	6.48	5.66	5.8	2.99	4.47	2.68
CoM std	2.74	1.76	2	1.52	1.74	1.76
	$M = 5.98$, $\Sigma = 0.44$, $\sigma = 2.21$			$M = 3.38$, $\Sigma = 0.96$, $\sigma = 1.68$		

8.2 Intra-Fraction Marker Motion Tracking in Fluoroscopic Image Sequences

After a description of a patient testing data set in Section 8.2.1, we report the results on intra-fraction marker motion tracking in Section 8.2.2.

8.2.1 Patient Fluoroscopic Data Set

The patient data set for intra-fraction motion tracking is the same as the set that was briefly described in Section 7.4.1 and used for some experiments on tracking validation. It consists of images of three prostate cancer patients. Patients 1, 2, and 3, respectively, have 10, 10, and 5 fluoroscopic image sequences acquired during the first three treatment fractions and once a week thereafter. The sequences were taken at 5.5 Hz for a duration of 30 seconds, which resulted in approximately 165 frames per sequence. Each sequence has a corresponding CBCT_e image obtained during the same treatment fraction that is used for marker localization in the first image of each sequence. The intra-fraction motion is measured as the displacement between the positions of the markers in the current image from their positions in the first image of the sequence.

8.2.2 Intra-Fraction Marker Motion

Figures 8.8 – 8.10 and their corresponding numerical values in Tables 8.6 – 8.8 provide marker motion statistics for the three patients in the AP/U and SI/V directions, and as the 2D displacements. The cumulative 2D intra-fraction motion statistics of the centre of mass of all markers and all patients combined is demonstrated in Figure 8.11. Analogous to inter-fraction analysis above, Table 8.9 provides intra-fraction motion error statistics, such as M, Σ, and σ, the computation of which was explained in Section 8.1.3.

The intra-fraction motion in our testing data set of three prostate patients was quite small: The 2D marker displacement was larger than 1 mm only in 2% of the images. The magnitude of the organ motion can be influenced by such things as dietary restrictions, immobilization equipment, patient positioning, protocols for treatment preparation and execution, and, finally, specific characteristics of individual patients. The intra-fraction displacements of the centre of mass of all markers in all patient image sequences were -0.04 ± 0.44 mm in the AP/U direction, 0.02 ± 0.3 mm in the SI/V direction, and 0.26 ± 0.26 mm in the 2D, respectively. This is smaller or similar to a number of other studies, where all mean and standard deviation values of the prostate intra-fraction motion were below 2 mm [129, 108, 163, 174, 8, 68]. While

the reported maximal shifts are typically well within 10 mm [68, 88, 16, 103, 163], some patients can display large motions, such as 16–25 mm [182, 103, 110] (see details in Section 5.3.1). As we allow the displacements of up to 20 mm in the AP/U and SI/V directions, and the rotation of up to 10° in-plane, the marker localization and tracking system proposed in this thesis, and implemented in the Gryphon software, is able to provide accurate and reliable tracking for different patients with a wide range of motion characteristics.

Figures 8.12 – 8.16 display the 2D trajectories of one of the three markers (M1, or the prostate base marker) in the UV coordinates in all fluoroscopic sequences of the three patients. The notation Fi means the i^{th} fluoroscopic image sequence of a patient. The positions in the first and last image of the sequence are denoted by the *circle* and *star*, respectively. Figure 8.17 shows the U and V motion in time of the same marker. While the marker trajectories of Patient 2 suggest clear correlation to the respiration motion, Patient 1 and 3 are less influenced by it (slightly noticeable in the V coordinate of Patient 1 and even less so in Patient 3), and are characterized by more erratic prostate motion. The fact that prostate motion may be affected by the respiratory cycle have been demonstrated in several studies, e.g., [103, 68].

This chapter concludes the part of the thesis devoted to method validation and case studies. Here, we have demonstrated the applicability of the methods proposed in this thesis to real patient images, and gave examples of typical statistical analysis that can be performed in the context of organ motion tracking and management in image-guided radiation therapy.

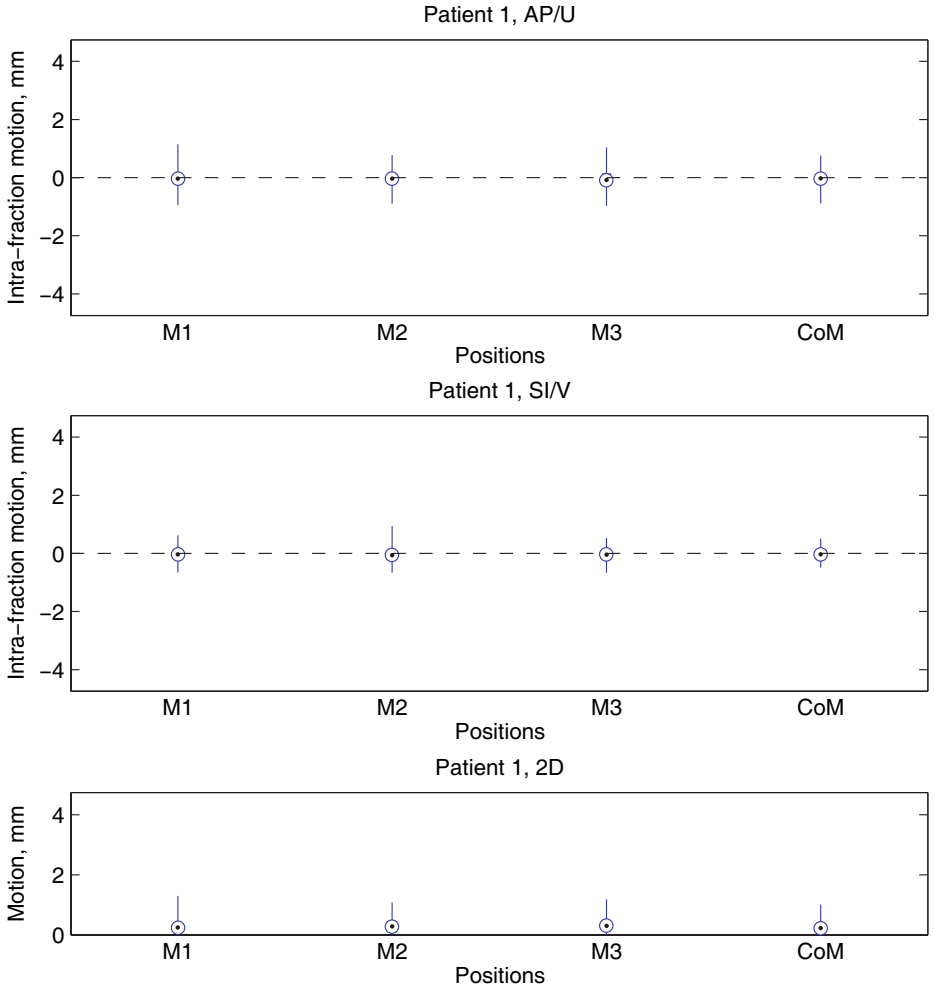


Figure 8.8: Intra-fraction motion for Patient 1.

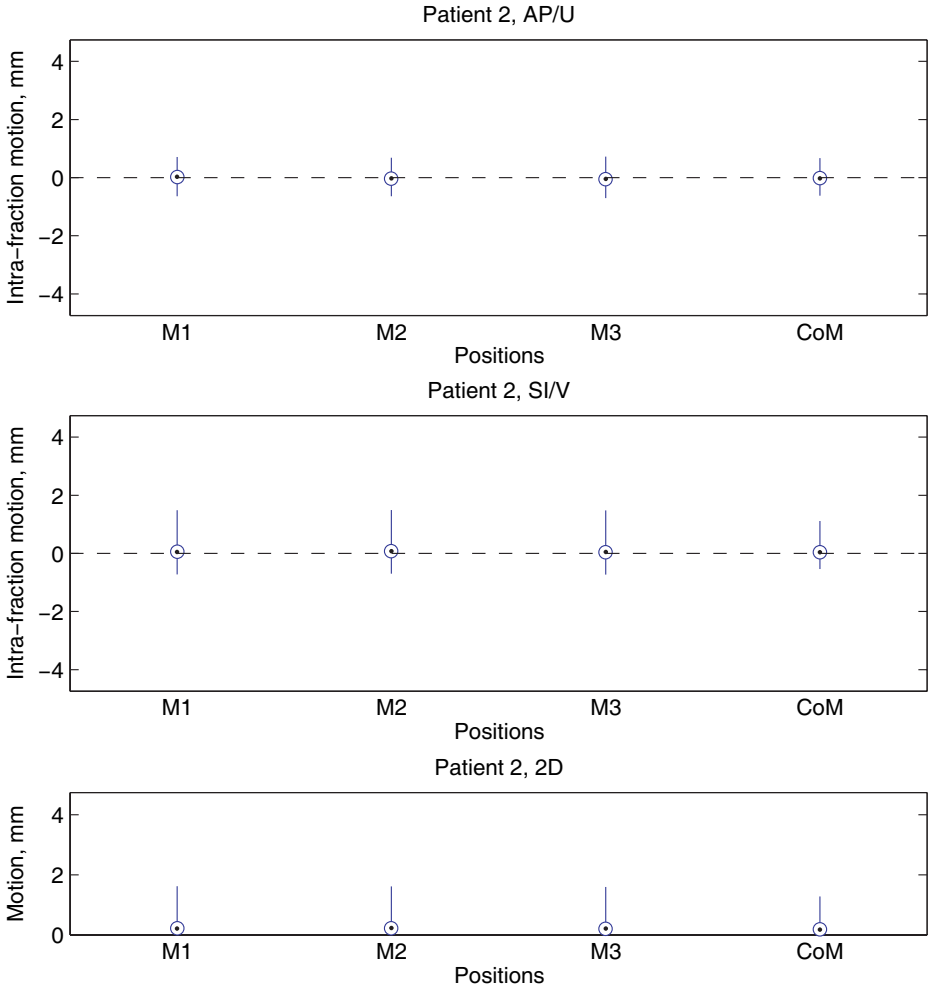


Figure 8.9: Intra-fraction motion for Patient 2.

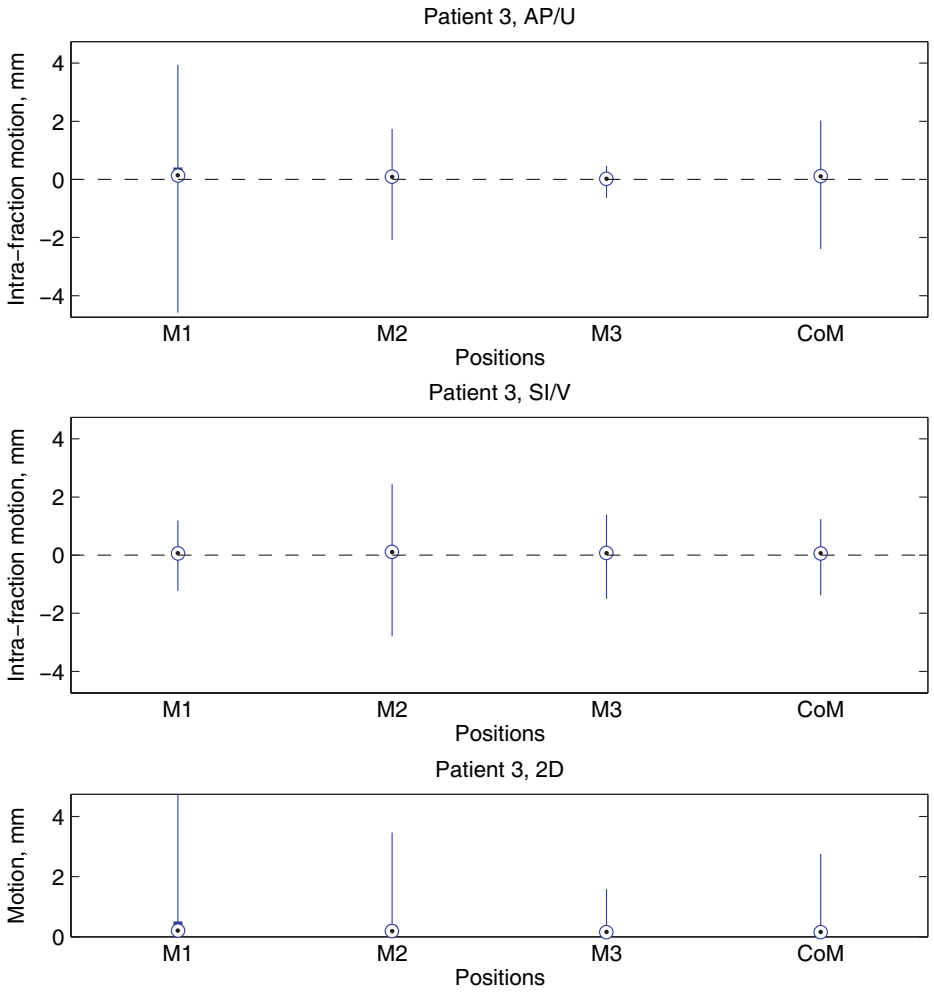


Figure 8.10: Intra-fraction motion for Patient 3.

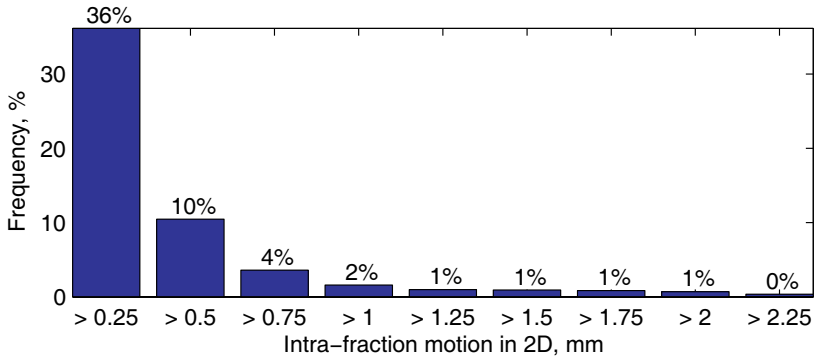


Figure 8.11: Cumulative 2D intra-fraction motion statistics for all patients.

Table 8.6: Intra-fraction motion for Patient 1.

	Intra-fraction motion, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
			Patient 1, AP/U			
M1	−0.94	−0.17	−0.03	0.12	1.14	-0.05 ± 0.33
M2	−0.90	−0.19	−0.03	0.14	0.78	-0.03 ± 0.28
M3	−0.97	−0.24	−0.08	0.15	1.04	-0.02 ± 0.34
CoM	−0.88	−0.18	−0.03	0.14	0.76	-0.03 ± 0.28
			Patient 1, SI/V			
M1	−0.65	−0.17	−0.03	0.11	0.62	-0.02 ± 0.22
M2	−0.66	−0.20	−0.06	0.10	0.94	-0.03 ± 0.25
M3	−0.67	−0.18	−0.04	0.11	0.53	-0.03 ± 0.21
CoM	−0.48	−0.13	−0.03	0.08	0.51	-0.02 ± 0.17
			Patient 1, 2D			
M1	0.01	0.15	0.25	0.42	1.30	0.32 ± 0.24
M2	0	0.17	0.28	0.44	1.08	0.32 ± 0.21
M3	0	0.18	0.30	0.44	1.18	0.34 ± 0.21
CoM	0.01	0.14	0.22	0.35	1.01	0.27 ± 0.19

Table 8.7: Intra-fraction motion for Patient 2.

	Intra-fraction motion, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
			Patient 2, AP/U			
M1	−0.64	−0.09	0.03	0.16	0.71	0.04 ± 0.22
M2	−0.64	−0.14	−0.03	0.09	0.69	−0.01 ± 0.2
M3	−0.70	−0.17	−0.05	0.08	0.72	−0.04 ± 0.2
CoM	−0.62	−0.13	−0.02	0.11	0.67	0 ± 0.2
			Patient 2, SI/V			
M1	−0.73	−0.07	0.05	0.22	1.48	0.12 ± 0.32
M2	−0.70	−0.06	0.08	0.26	1.49	0.14 ± 0.33
M3	−0.73	−0.08	0.05	0.21	1.48	0.1 ± 0.32
CoM	−0.54	−0.05	0.04	0.18	1.11	0.09 ± 0.24
			Patient 2, 2D			
M1	0	0.11	0.21	0.35	1.62	0.26 ± 0.2
M2	0.01	0.13	0.22	0.34	1.62	0.27 ± 0.19
M3	0.01	0.13	0.21	0.31	1.60	0.25 ± 0.18
CoM	0.01	0.11	0.18	0.28	1.28	0.21 ± 0.15

Table 8.8: Intra-fraction motion for Patient 3.

	Intra-fraction motion, mm					
Position	Min	25%	Med	75%	Max	Mean ± Std
			Patient 3, AP/U			
M1	−4.58	−0.02	0.14	0.41	3.94	-0.25 ± 1.66
M2	−2.09	−0.04	0.09	0.20	1.73	-0.1 ± 0.73
M3	−0.63	−0.08	0.02	0.11	0.47	-0.01 ± 0.19
CoM	−2.39	−0.04	0.11	0.21	2.03	-0.12 ± 0.85
			Patient 3, SI/V			
M1	−1.23	−0.03	0.07	0.16	1.19	0.01 ± 0.48
M2	−2.79	−0.01	0.11	0.28	2.44	-0.08 ± 1.04
M3	−1.50	−0.03	0.07	0.17	1.40	-0.01 ± 0.57
CoM	−1.38	−0.02	0.06	0.15	1.24	-0.02 ± 0.52
			Patient 3, 2D			
M1	0	0.10	0.20	0.50	4.74	0.64 ± 1.01
M2	0	0.10	0.18	0.40	3.47	0.5 ± 0.74
M3	0	0.09	0.15	0.28	1.58	0.3 ± 0.35
CoM	0	0.09	0.15	0.31	2.76	0.39 ± 0.58

Table 8.9: Intra-fraction error statistics.

	Intra-fraction error statistics, mm		
Position	Patient 1	Patient 2	Patient 3
AP/U			
CoM mean	−0.03	0	−0.12
CoM std	0.28	0.2	0.85
	$M = -0.05$, $\Sigma = 0.06$, $\sigma = 0.53$		
SI/V			
CoM mean	−0.02	0.09	−0.02
CoM std	0.17	0.24	0.52
	$M = 0.02$, $\Sigma = 0.06$, $\sigma = 0.34$		
2D			
CoM mean	0.27	0.21	0.39
CoM std	0.19	0.15	0.58
	$M = 0.29$, $\Sigma = 0.09$, $\sigma = 0.36$		

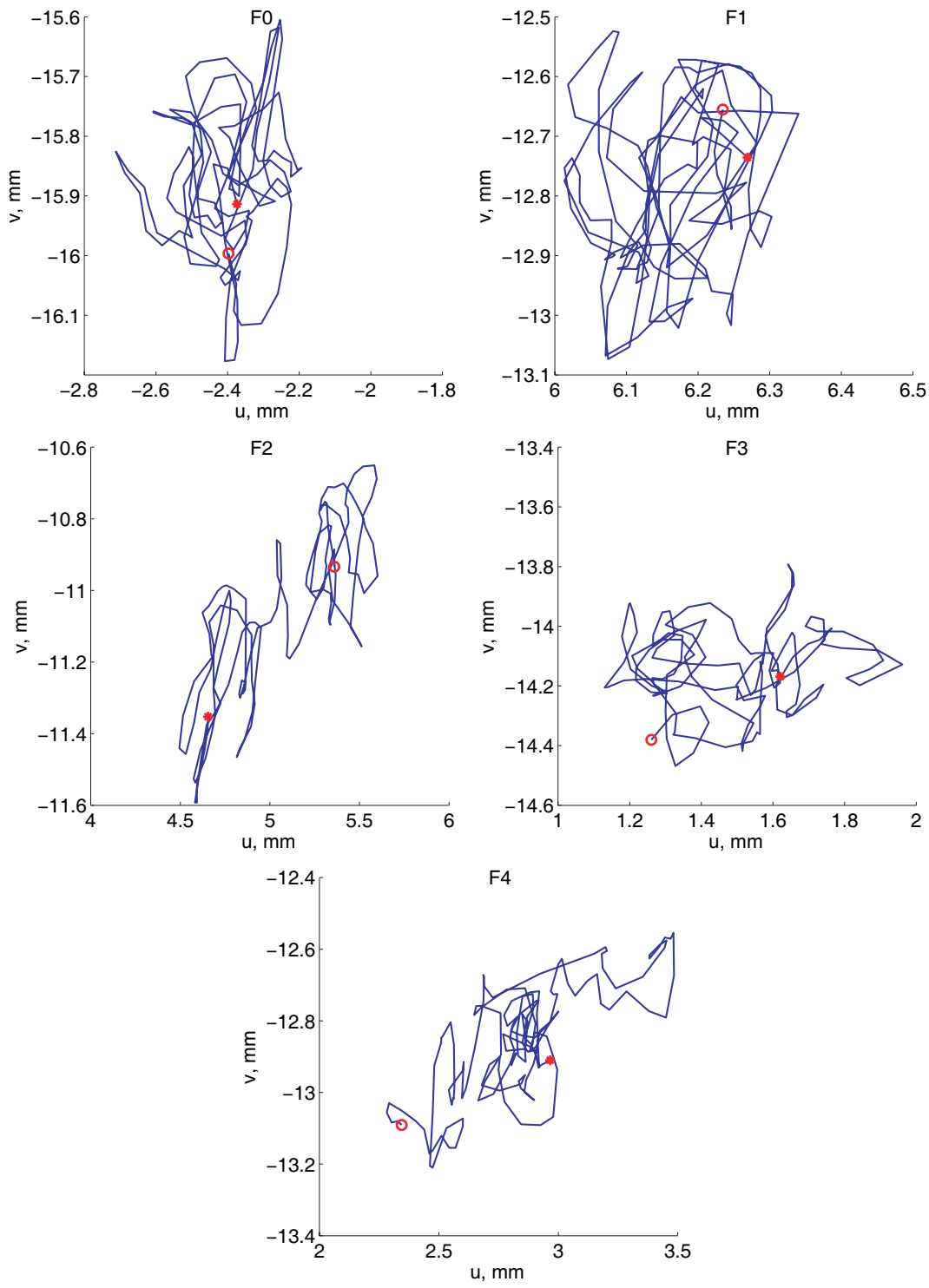


Figure 8.12: Two-dimensional trajectories of the markers in fluoroscopic sequences F0 to F4 of Patient 1.

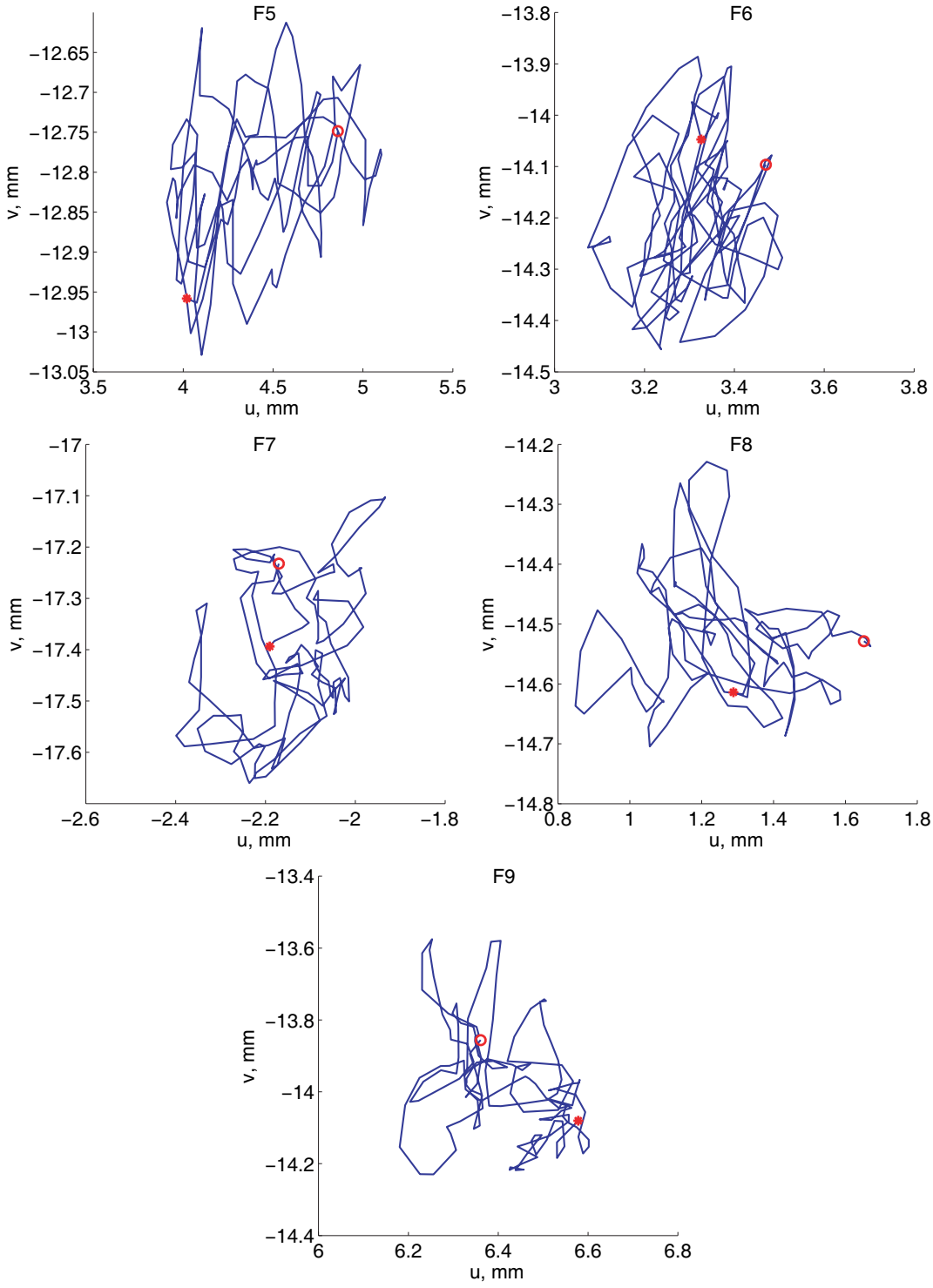


Figure 8.13: Two-dimensional trajectories of the markers in fluoroscopic sequences F5 to F9 of Patient 1.

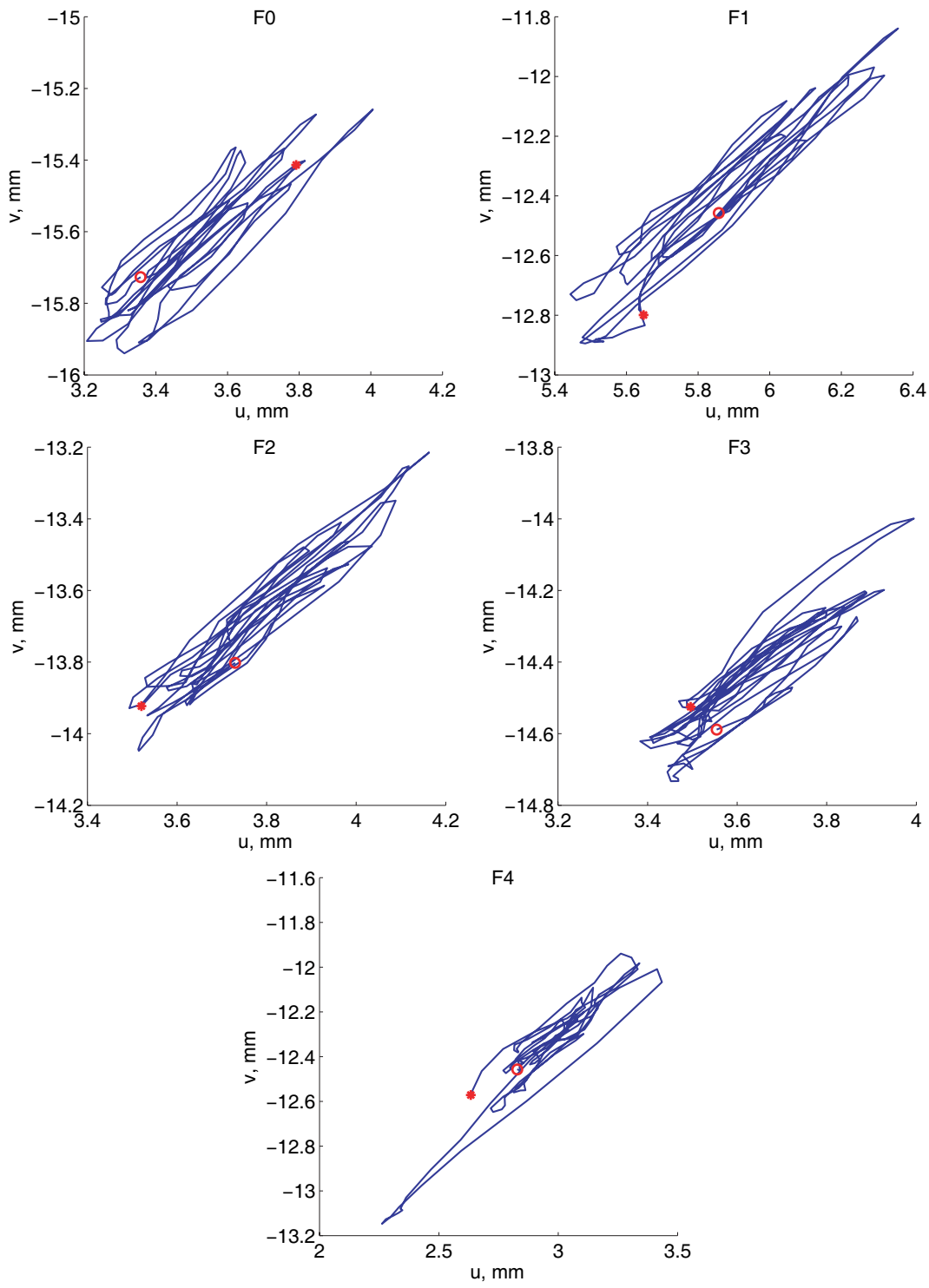


Figure 8.14: Two-dimensional trajectories of the markers in fluoroscopic sequences F0 to F4 of Patient 2.

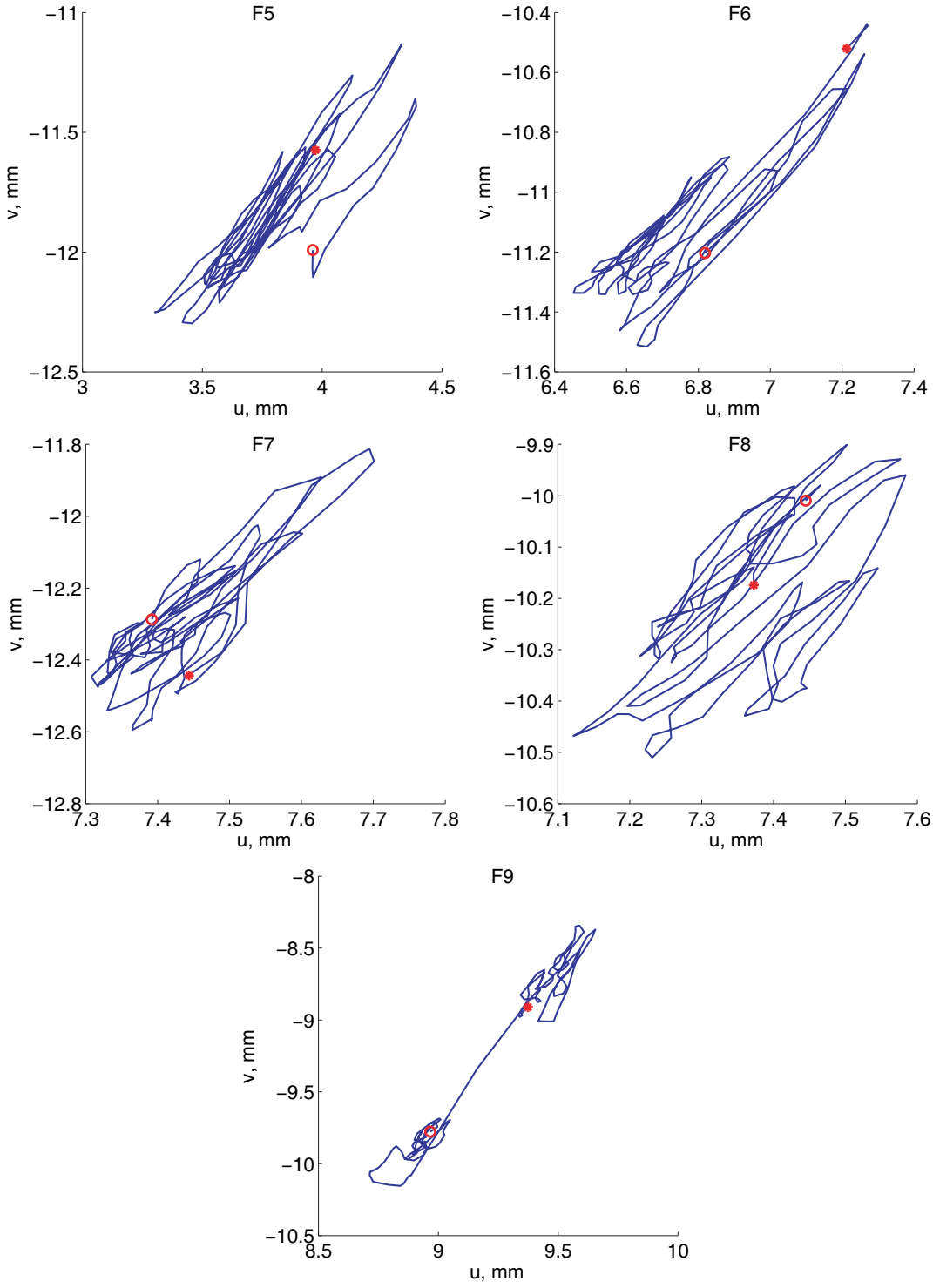


Figure 8.15: Two-dimensional trajectories of the markers in fluoroscopic sequences F5 to F9 of Patient 2.

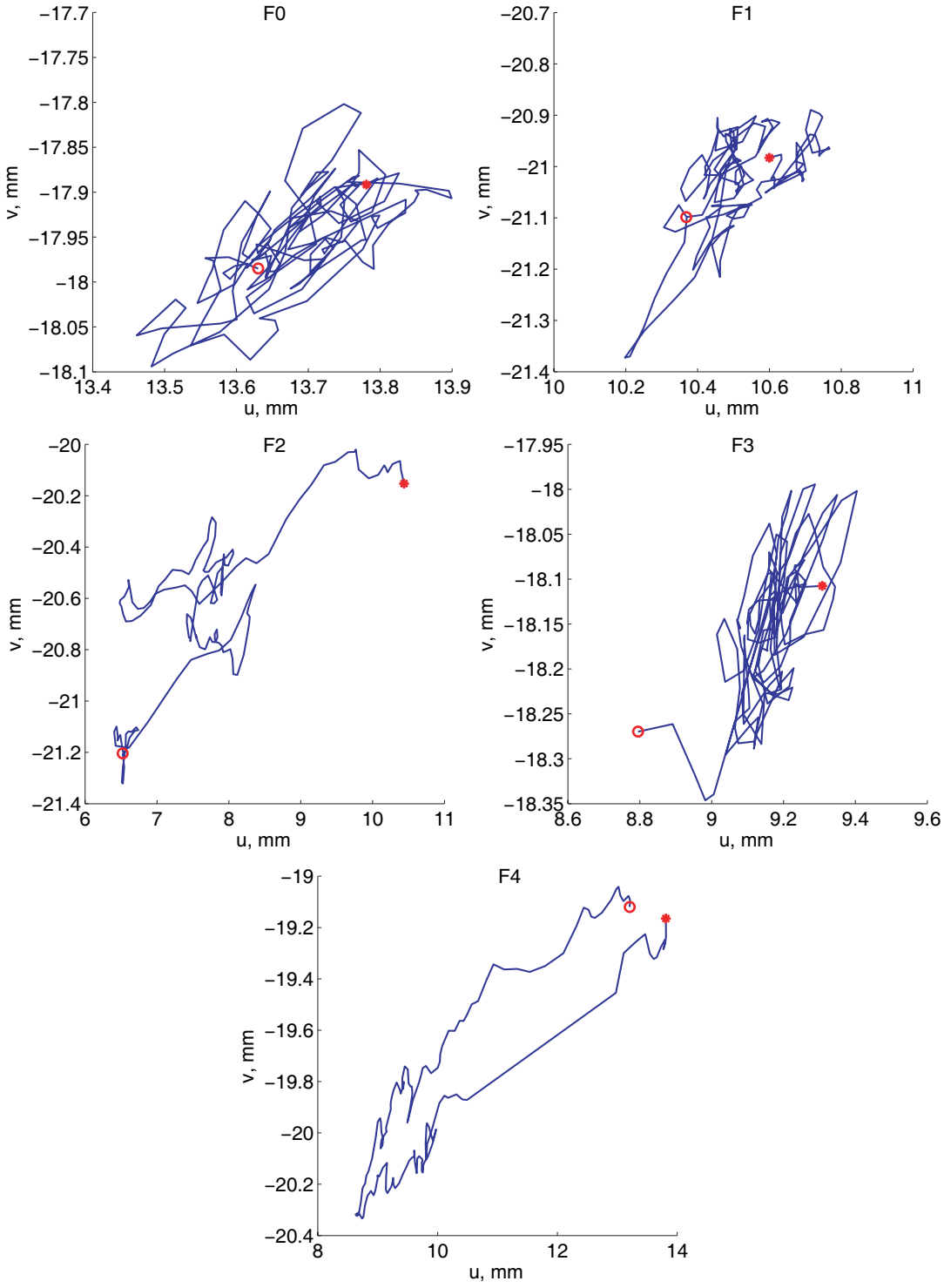


Figure 8.16: Two-dimensional trajectories of the markers in fluoroscopic sequences F0 to F4 of Patient 3.

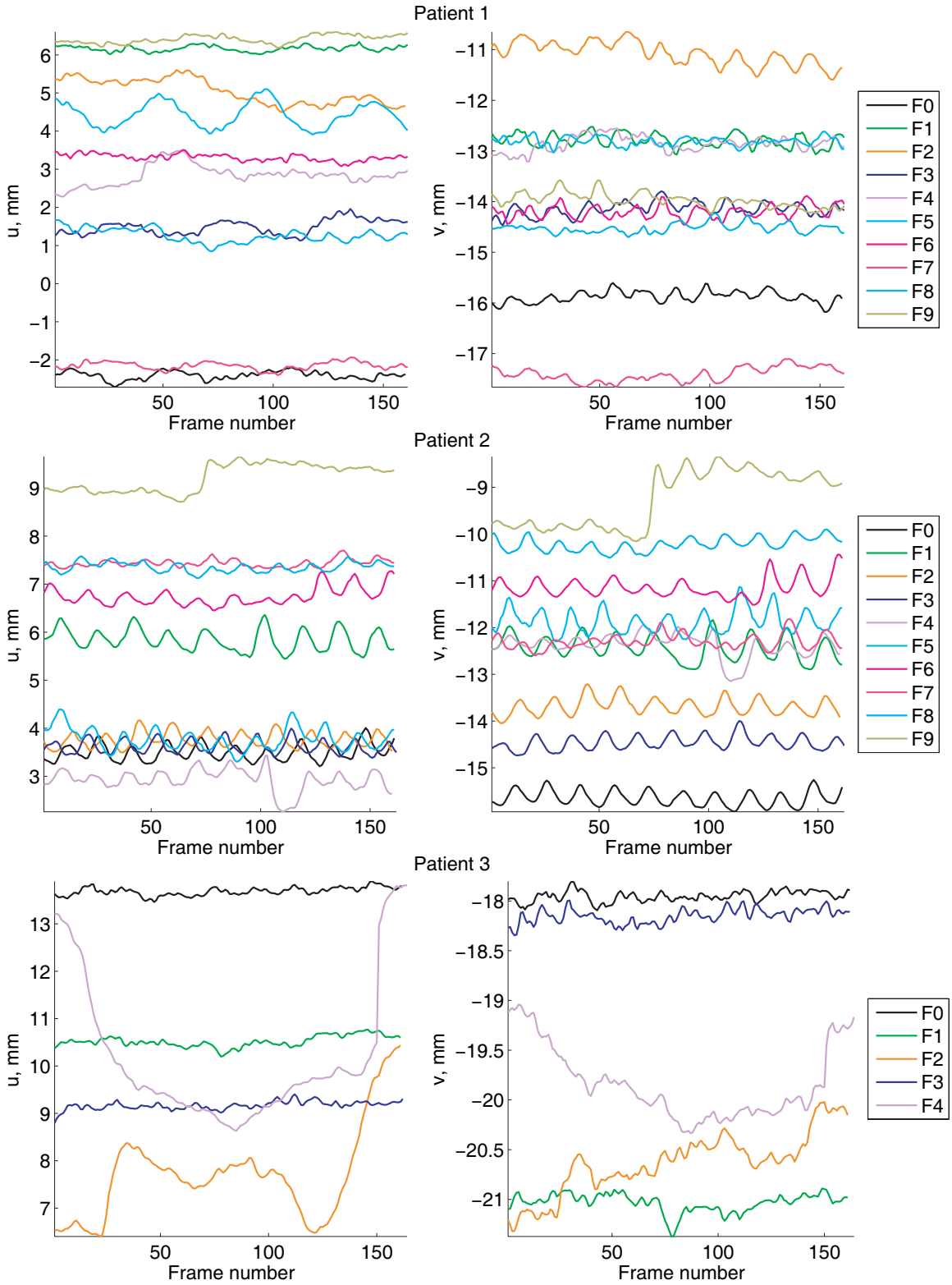


Figure 8.17: Intra-fraction marker motion in time for Patient 1 (*top row*), 2 (*middle row*), and 3 (*bottom row*).

Chapter 9

Conclusions and Further Directions

In this chapter, we summarize the contributions of this thesis in Section 9.1, and discuss possible extensions and directions for future work in Section 9.2.

9.1 Summary and Contributions

This thesis described the design and development of a complete system for automatic marker localization and tracking. After the problem background information in Chapter 1, the requirements, assumptions, and overview of the proposed approach were described in Chapter 2. Then, Chapters 3 – 6 provided modelling details for each of the four parts of the system, in particular, 3D marker segmentation and modelling, filtering procedures, 2D marker localization, and tracking. The algorithms developed in these chapters were extensively validated in Chapter 7. Chapter 8 demonstrated the application of the proposed system to patient images. Some of the modelling, formulation, and implementation details have been assembled in Appendices A – D. This thesis contains a number of contributions that we describe below. For convenience, we group them under the topics of (i) mathematical modelling, (ii) system design and selection of the components, and (iii) implementation and computational efficiency.

Mathematical modelling

One of the initial ideas proposed in this thesis, on which the approach for marker localization and tracking in fluoroscopic sequences was successfully constructed, was to use a higher-contrast 3D CBCT image acquired during the same treatment fraction in order to assist in initialization of the positions of the markers in the lower quality fluoroscopic images. First, the markers were localized in the 3D CBCT. Although this required a multi-step segmentation algorithm followed by the correction procedures, it was still easier to localize the markers in the 3D compared to the 2D x-ray projection

images, where multiple anatomical features were overlayed, resulting in low contrast of the markers. Then, the 3D marker positions, knowledge of the system geometry, and assumptions about motion that could have happened between the acquisition of the CBCT and fluoroscopic images, were used to define a registration region of interest (ROI). Further, the 2D-3D image registration between the fluoroscopic and CBCT images was performed on this ROI. The transformation found as a result of the registration incorporated the displacement of the target between the acquisition of the images, and possible calibration inaccuracies. Finally, the 2D marker positions in the first image of the fluoroscopic sequence were computed by application of the found transformation to the 3D marker positions.

A common approach for registration of the 2D x-ray and 3D CBCT images by using a digitally reconstructed radiograph (DRR) [42, 154, 249] could not provide satisfying results in the proposed application. First, the DRR generation typically involves significant computational resources. And second, our goal was to conduct the registration based on the markers rather than any other features that can be found in the CBCT and fluoroscopic images. To address these issues, a registration framework was proposed, in which the fluoroscopic images were preprocessed by a specially designed *marker enhancement filter* (MEF) that amplified the markers, and suppressed the background and anatomical features. At the same time, instead of generating a complete DRR from the CBCT image, a 2D template image was created using only the markers localized in the 3D. To distinguish this new imaging modality, a new term was introduced – a *selective DRR*. Hence, the proposed registration framework creates two images that contain measurements of the markers with very similar intensity characteristics. In addition to computational savings related to DRR generation, and the ability to register images based on the markers, such approach also paved the way for the use of registration tools more typical of the single-modality registration, which are simpler, more reliable, and more computationally efficient. In particular, the proposed framework made it possible to use a rigid geometric transformation, as we consider the prostate to be a rigid organ for the purposes of marker localization and tracking (Section 2.2). In contrast, registration of the complete images containing the prostate with its surrounding soft tissues may have required a non-rigid registration method.

A similar approach was undertaken in tracking, where the template image was created from the markers localized in the first image of the sequence by one of the proposed image generation methods described below, and all subsequent images undergoing 2D-2D registration with the template were preprocessed by a simplified version of the MEF.

The following novel mathematical modelling techniques were proposed within

the registration framework discussed above:

- A definition of the 3D *marker set* was formalized as a set of points within an image (voxel centres in a discrete image) possessing high intensities and located in a close geometrical proximity that correspond to a measurement of a fiducial marker. The 3D marker sets were formed as a result of the segmentation procedure described in Section 3.1, consisting of the thresholding to select high-intensity points, clustering to group the points by their location proximity, and checking the marker size selection criteria. Finally, the candidate marker sets obtained by segmentation were analyzed by the correction procedures explained in Section 3.3 that were derived based on our knowledge about the proposed application. In addition to careful selection and combination of the existing components to successfully solve the problem, a novel *adaptive thresholding technique* was designed, which selects high-intensity points by exploiting information about the number and size of the fiducial markers (Section 3.1.1). As the proposed method does not rely on particular intensity threshold or range, it is applicable to the 3D images possessing different intensity ranges, such as the case of the CBCT images reconstructed with different voxel sizes described in the introduction to Chapter 3.
- The *region of interest* (ROI), on which the registration procedures of both marker localization and tracking steps are to be performed, may greatly influence the registration outcome by affecting both reliability and computational time. The usage of a smaller ROI within the image, centered around predicted marker positions, proved to be an efficient approach in a number of studies [141, 204, 7, 169]. The ROI was either selected manually or as a circular or square area of the predefined size. Although the application of the fixed ROI size based on the assumptions of the target motion magnitude may seem attractive due to its simplicity, it may easily become a trigger for failure in the software implementation, as in our case the range of allowable target motion is defined in terms of both shift and rotation. In addition, the experiments were conducted using both 2D and 3D rigid transformations. While it is desirable to select the ROI as small as possible to decrease computational time and improve reliability of the subsequent image registration, a ROI that is too small may result in the markers moving outside of the ROI while transformed by the legitimate parameter values. To solve this problem, the approach of adaptive computing of the ROI size based on the geometric transformation parameters was proposed. The details are provided in Section 4.3.
- While the registration ROI was computed from the positions of the 3D marker sets and knowledge of the system geometry and calibration, the precise 2D

marker positions in the fluoroscopic image were recovered by applying the 2D-3D image registration. The voxels in the CBCT may often be of a larger size than the pixels in the fluoroscopic image, which may result in larger 3D markers compared to their counterparts in the 2D. This can potentially increase geometric uncertainty during registration. In addition, it is not uncommon for the CBCT images to have streak artifacts around the markers, i.e., a number of bright voxels surrounding the high-contrast marker, which do not represent a real measurement, and are not present in the corresponding fluoroscopic images. To overcome these issues, it was proposed to estimate the *marker models* from the 3D markers sets (Section 3.4). The models are mathematical representations of the markers in 3D, such as continuous cylindrical or spherical models of the known dimensions. Detailed formulations are provided in Appendix C.3. As was discussed in Section 3.5, the models of the fiducial markers, which are 3 mm long and 0.8 mm in diameter, can be successfully computed from the 3D marker sets segmented in the CBCT images with the voxel sides of 0.25 and 0.5 mm, and in most CBCTs with 1 mm voxels. For the larger voxel sizes, the cylinder orientations could not be reliably recovered for such a small marker size.

- To address different CBCT resolutions and quality, three methods were proposed for *selective DRR* generation that operated on the 3D marker models, marker sets, and marker positions, respectively. The *marker model DRR* simulates an x-ray image creation process by applying a raytracing principle (Section 5.1.1). It is a good option for the CBCT images, for which the 3D marker models can be reliably estimated, as it creates the 2D markers that are very similar in shape to the ones found in the fluoroscopic images. An alternative that is computationally cheaper and is also suitable for the markers of irregular shapes, such as long flexible coils [7], is the *voxel splat DRR* (Section 5.1.2). This method "projects" the voxels belonging to the 3D marker sets onto the 2D plane. It incorporates the notion of the voxel size, hence it can create contiguous features in the 2D from the contiguous features in the 3D, regardless of a difference in size between the CBCT voxels and fluoroscopic pixels. For the CBCTs of lower quality and resolution, a *marker splat DRR* method was proposed (Section 5.1.3), for which no marker orientation information is used. The markers in the generated image are represented by round spots of the size comparable to the fiducial marker. Being very computationally efficient, this method may also be more suitable for applications, for which execution time is important, such as in real-time motion tracking. Finally, in the 2D-2D image registration of the tracking procedure, the template was generated from the markers localized in the first image of the

sequence by using a 2D version of the marker splat DRR, a *2D marker splat image* (Section 6.1).

- To preprocess fluoroscopic images prior to image registration, a novel *marker enhancement filter* (MEF) was proposed, as no existing filter that was considered could provide satisfying results for our application. The proposed filter was designed to amplify the features of a certain size by conducting local intensity analysis. In brief, the MEF is built upon two images, where the first one amplifies the regions where the markers can be found with high probability, while the second one smoothes and preserves marker shapes found within these regions. The MEF also includes a novel *contrast enhancement function*. The details of the MEF design were provided in Section 4.4.

  A simplified version of the MEF, called the *magnitude-only filter* (MOF), was proposed for use in tracking (Section 4.4.5). It only amplifies the marker shapes without the search for the probable marker regions, as we assume that the displacement between the markers in the subsequent images of the sequence is rather small, and hence the regions can be formed around the marker positions found in the preceding image. The use of the MOF resulted in higher computational efficiency of the tracking compared to application of the full MEF (5-fold decrease in time in our Matlab implementation) or the conventional 2D-2D image registration approaches that use interpolation (7–100-fold improvement). For details, see Sections 6.1 and 7.4.2.

System design and selection of the components

This thesis includes multiple literature reviews, justifications, and experiments to select the components that were build into a complete marker localization and tracking system:

- Prior to performing any processing on fluoroscopic images, including the MEF or MOF application, it was desirable to reduce the noise. Common spatial filters could not provide sufficient improvement without oversmoothing the markers. Instead, *temporal filtering* was used (Section 4.1). Several existing temporal filters used in fluoroscopy were evaluated, such as a simple form of recursive filtering, a filter based on the Karhunen-Loève transform (KLT), and finite impulse response filters, of which, to the best of our knowledge, only the mean filter (frame averaging) was reported to be used in fluoroscopy. An experiment with the images of the Quasar motion phantom demonstrated that the Dolph-Chebyshev filter provided the best improvement in the contrast-to-noise ratio in presence of the moving markers, and significantly outperformed the mean filter.
- The *design of the MEF* was motivated by reviewing and evaluating a number

of existing methods, such as morphological opening, template matching, the marker extraction kernel (Section 4.2.1), and multiple filters exploiting the scale-space approach that were mostly designed for vessel detection and computer vision applications (Section 4.2.2). None of these filters provided satisfying results. Therefore, the MEF, and its simplified version, the MOF, were designed as described in Sections 4.4.1 and 4.4.5.

- The *optimization method* to solve the registration problem was selected among the existing approaches based on computational cost of our implementation (the methods provided similar accuracy). However, the selection of an appropriate *starting point*, which is crucial for successful registration using the functions seeking a local optimum, required extensive evaluation (Section 5.3.3). Based on our knowledge about the problem, multiresolution and exhaustive search over sampled parameter space, and their modifications and combinations, were considered as the most promising approaches. The use of conventional multiresolution approach, in which the registration is performed on the highest level first (the smallest, most blurred images), and then the solution is improved for images with the increasing resolution, have not resulted in 100% localization success rates for most data sets, unlike the other evaluated methods. The exhaustive search is often implemented as the normalized cross-correlation between the image and the template, performed over one pixel shifts in 2D. To reduce the computation, and incorporate rotation in addition to the shifts, the method was modified by adding multiple templates with different rotation angles and performing the cross-correlation on reduced resolution.
- The appropriate *distance measure* for the image registration problem was selected by evaluating existing functions on our sample images (Section 5.2).
- *Bounds* on geometric transformation parameters were selected by reviewing multiple studies to determine average and maximal displacements observed in prostate patients (Section 5.3.1).
- Based on knowledge of the system geometry, marker placement, image resolution, and expected magnitude of the marker displacement, it was determined that the proposed imaging system (including hardware, geometric setup, and image acquisition software) was not sensitive enough to recover the out-of-plane marker displacements with reasonable accuracy. Therefore, it was decided to use the *2D geometric transformation* in our registration problems (Section 5.3.2). This justification was also reinforced by the experiments with the phantom images (Section 7.3.3).
- Due to the fact that the displacement between the markers in the neighbouring images of the sequence is small, and, therefore, the registration ROI can be tight,

it was speculated that it may be simpler and more computationally efficient to conduct 2D-2D image registration in a conventional way, i.e., to register images directly with the use of *interpolation*, without the application of the MOF and the 2D marker splat image generation method (Section 6.1). The experiment showed that while both approaches provided acceptable registration accuracy, the application of the MOF and the 2D marker splat was, in fact, much cheaper computationally (7–100 times, Section 7.4.2). The components for the image registration problem used in tracking, such as the distance measure, optimization method, optimization starting point, and geometric transformation were selected based on our findings for the 2D-3D image registration problem used in marker localization (Sections 6.1 and 6.3).

- By analyzing the literature on organ motion tracking, several promising *prediction models* were selected for evaluation (Section 6.2). For the data set containing multiple sequences of the Quasar motion phantom, where markers perform different kinds, magnitudes, and speeds of motion, it was concluded that the linear regression model based on three preceding images was the most suitable predictor for the proposed tracking procedure.

Implementation and computational efficiency

Even for post-processing applications, such as those described in this thesis, computational efficiency is important, and paves the way for the future real-time application of the proposed methods. When conducting experiments to select the methods and their parameters (Chapter 7), the algorithms were verified to provide expected results, reliable operation, and low computational cost. For example, the images were quite large, and in many cases it would not be possible to operate on the whole image or to have several images in memory simultaneously. Therefore, once the image was opened, a ROI was typically selected and kept in memory, and all further processing was performed on this ROI. This reduced both memory use, and execution time of the computations. Another example is temporal filtering, where the images were opened one at a time, multiplied by the corresponding filter coefficients, and added to the resulting filtered image. Also, prior to template image generation, a ROI was computed for each of the future generated markers first, and then computation was conducted only in those ROIs.

As for the implementation, we described algorithms for every part of the system and explained implementation details where appropriate and necessary in Appendix D.

9.2 Extensions and Future Work

For the methods proposed in this thesis, the following possible extensions and future developments are suggested:

- Further collaboration with clinicians may result in the selection of the projects and applications to which the methods proposed in this thesis can be applied directly or with modifications. For example, the current design of our system can be used to track markers in the CBCT projections, where the angle of the imaging system changes for each image of the sequence. Such tracking routine can be used for 3D motion monitoring during the treatment. As an example, our methods can be used to improve the tracking success of the kV monitoring systems such as the one described by Ng *et al.* [164]. Another example is the use of the MEF for marker segmentation in the stereotactic body radiation therapy tracking system [15].
- Further evaluation of the marker localization and tracking approach proposed in this thesis in application to fluoroscopic images produced using different settings of the kV imaging system may result in finding an opportunity for reduction of imaging dose to the patients.
- The methods proposed in this thesis were evaluated by implementing them as a Matlab software module. This was appropriate for selection of the methods and their components, proof-of-concept demonstration, and post-processing of the data. For the clinical use, however, it may be more reasonable to implement the algorithms in one of the more efficient software environments using C++, C#, or Java, which may eventually enable the use of the proposed algorithms in real-time applications.
- In the context of real-time tracking, it is also possible to modify the proposed system to combine external motion tracking with occasional x-ray imaging for certain anatomical sites, similar to [193, 95, 44, 240].
- The fluoroscopic image sequences used in the experiments described in this thesis were acquired within the treatment fraction when the MV irradiation was turned off. The MV scatter further reduces the contrast of the features and increases the noise. The methods proposed in this thesis can be further evaluated on the data sets acquired simultaneously with the MV irradiation. It is expected that our tracking system may still work reasonably well: For example, the MEF showed promising preliminary results in application to the portal images (Section 7.2.4), which possessed significantly lower contrast. If successful, and implemented for the real-time use, our proposed methods can be applied to organ motion monitoring simultaneously with the MV radiation

delivery. Encouraging results were reported by Arumugam *et al.* [15], who used the MEF for marker segmentation in their tracking system that showed similar performance for both with and without simultaneous MV irradiation.

- In addition to testing with the MV beam turned on, it may be interesting to evaluate the performance of the MEF and the whole proposed marker localization and tracking system with different imaging modalities, such as portal or ultra-sound images, which may create opportunities to use the proposed methods in more applications.
- The ease of use of the automatic tracking system may eventually lead to the possibility of designing custom and adaptive treatment plan for each patient. In such a system, the motion patterns of each specific patient can be used to determine treatment parameters, and those parameters can be adapted over time should the motion patterns change.
- Finally, though this thesis presents the methods for marker localization and tracking, modifications of the system can be used for markerless tracking in gating and tumour tracking applications. Instead of the markers, other features can be detected and tracked, such as edges, blobs, and intensity changes between high and low x-ray attenuation regions. While the components of the system will have to be adjusted, the general framework can be reused.

Bibliography

[1] *Matlab Documentation*, Mathworks, http://www.mathworks.com/help/matlab, R2009b.

[2] *Multileaf Collimators*, Varian Medical Systems, http://newsroom.varian.com, Image Gallery section.

[3] *University Health Network Research Ethics Board (UHN-REB)*, Toronto, ON, http://www.uhnres.utoronto.ca/reb.

[4] *ICRU Report 50: Prescribing, recording, and reporting photon beam therapy*, Bethesda, MD, 1993.

[5] *Radiotherapy equipment – coordinates, movements and scales, IEC-61217*, 1996.

[6] *ICRU Report 62: Prescribing, recording, and reporting photon beam therapy (Supplement to ICRU Report 50)*, Bethesda, MD, 1999.

[7] J. Adamson and Q. Wu, *Prostate intrafraction motion evaluation using kV fluoroscopy during treatment delivery: A feasibility and accuracy study*, Medical Physics, 35 (2008), pp. 1793–806.

[8] ——, *Inferences about prostate intrafraction motion from pre- and posttreatment volumetric imaging*, International Journal of Radiation Oncology, Biology, Physics, 75 (2009), pp. 260–7.

[9] N. M. Alpert, J. F. Bradshaw, D. Kennedy, and J. A. Correia, *The principal axes transformation – A method for image registration*, The Journal of Nuclear Medicine, 31 (1990), pp. 1717–22.

[10] L. E. Antonuk, *Electronic portal imaging devices: A review and historical perspective of contemporary technologies and research*, Physics in Medicine and Biology, 47 (2002), pp. R31–65.

[11] L. E. Antonuk, J. Boudry, W. Huang, D. L. McShan, E. J. Morton, J. Yorkston, M. J. Longo, and R. A. Street, *Demonstration of megavoltage and diagnostic x-ray imaging with hydrogenated amorphous silicon arrays*, Medical Physics, 19 (1992), pp. 1455–66.

[12] S. Aouadi and L. Sarry, *Evaluation of a stochastic approach to 2D-3D intensity-based registration*, in EMBS, 29th Annual International Conference of the IEEE, Lyon, France, 2007, IEEE, pp. 6387–90.

[13] S. Aouadi and L. Sarry, *Accurate and precise 2D-3D registration based on X-ray intensity*, Computer Vision and Image Understanding, 110 (2008), pp. 134–51.

[14] S. Arslan, A. Yilmaz, B. Bayramgürler, Ö. Uzman, E. Ünver, and E. Akkaya, *CT-guided transthoracic fine needle aspiration of pulmonary lesions: Accuracy and complications in 294 patients*, Medical Science Monitor, 8 (2002), pp. CR493–7.

[15] S. Arumugam, M. Sidhom, A. Xing, and L. Holloway, *An online x-ray based position validation system for prostate hypofractionated radiotherapy*, Medical Physics, 43 (2016), pp. 961–74.

[16] J.-F. Aubry, L. Beaulieu, L.-M. Girouard, S. Aubin, D. Tremblay, J. Laverdière, and E. Vigneault, *Measurements of intrafraction motion and interfraction and intrafraction rotation of prostate by three-dimensional analysis of daily portal imaging with radiopaque markers*, International Journal of Radiation Oncology, Biology, Physics, 60 (2004), pp. 30–9.

[17] R. Aufrichtig and D. L. Wilson, *X-ray fluoroscopy spatio-temporal filtering with object detection*, IEEE Transactions on Medical Imaging, 14 (1995), pp. 733–46.

[18] M. Bal and L. Spies, *Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering*, Medical Physics, 33 (2006), pp. 2852–9.

[19] I. N. Bankman, ed., *Handbook of Medical Imaging, Processing and Analysis*, Academic Press, San Diego, CA, 2000.

[20] M. Baumann, P. Mozer, V. Daanen, and J. Troccaz, *Transrectal ultrasound prostate biopsy tracking with efficient and accurate deformation estimation*, in IEEE ISBI: From Nano to Macro, IEEE, 2009, pp. 1211–4.

[21] A. Bayley, T. Rosewall, T. Craig, R. Bristow, P. Chung, M. Gospodarowicz, C. Ménard, M. Milosevic, P. Warde, and C. Catton, *Clinical application of high-dose, image-guided intensity-modulated radiotherapy in high-risk prostate cancer*, International Journal of Radiation Oncology, Biology, Physics, 77 (2010), pp. 477–83.

[22] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, SIAM, Philadelphia, PA, 2001.

[23] R. I. Berbeco, S. B. Jiang, G. C. Sharp, G. T. Y. Chen, H. Mostafavi, and H. Shirato, *Integrated radiotherapy imaging system (IRIS): Design considerations of tumour tracking with linac gantry-mounted diagnostic x-ray systems with flat-panel detectors*, Physics in Medicine and Biology, 49 (2004), pp. 243–55.

[24] R. I. Berbeco, H. Mostafavi, G. C. Sharp, and S. B. Jiang, *Towards fluoroscopic respiratory gating for lung tumours without radiopaque markers*, Physics in Medicine and Biology, 50 (2005), pp. 4481–90.

[25] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1999.

[26] P. Bifulco, M. Cesarelli, R. Allen, M. Romano, A. Fratini, and G. Pasquariello, *2D-3D registration of CT vertebra volume to fluoroscopy projection: A calibration model assessment*, EURASIP Journal on Advances in Signal Processing, 2010 (2010), pp. 1–8.

[27] W. Birkfellner, M. Figl, J. Kettenbach, J. Hummel, P. Homolka, R. Schernthaner, T. Nau, and H. Bergmann, *Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic CT images*, Medical Physics, 34 (2007), pp. 246–55.

[28] W. Birkfellner, R. Seemann, M. Figl, J. Hummel, C. Ede, P. Homolka, X. Yang, P. Niederer, and H. Bergmann, *Wobbled splatting – A fast perspective volume rendering method for simulation of x-ray images from CT*, Physics in Medicine and Biology, 50 (2005), pp. N73–84.

[29] W. Birkfellner, J. Wirth, W. Burgstaller, B. Baumann, H. Staedele, B. Hammer, N. C. Gellrich, A. L. Jacob, P. Regazzoni, and P. Messmer, *A faster method for 3D/2D medical image registration – A simulation study*, Physics in Medicine and Biology, 48 (2003), pp. 2665–79.

[30] J.-P. Bissonnette, D. Moseley, E. White, M. Sharpe, T. Purdie, and D. A. Jaffray, *Quality assurance for the geometric accuracy of cone-beam CT guidance in radiation therapy*, International Journal of Radiation Oncology, Biology, Physics, 71 (2008), pp. S57–61.

[31] G. J. Bootsma, J. H. Siewerdsen, M. J. Daly, and D. A. Jaffray, *Initial investigation of an automatic registration algorithm for surgical navigation*, in EMBS, 30th Annual International Conference of the IEEE, Vancouver, BC, 2008, IEEE, pp. 3638–42.

[32] N. K. Bose, *Digital Filters*, North-Holland, 1985.

[33] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK; New York, 2004.

[34] L. G. Brown, *A survey of image registration techniques*, ACM Computing Surveys, 24 (1992), pp. 325–76.

[35] D. Buck, M. Alber, and F. Nüsslin, *Potential and limitations of the automatic detection of fiducial markers using an amorphous silicon flat-panel imager*, Physics in Medicine and Biology, 48 (2003), pp. 763–74.

[36] T. Budiharto, P. Slagmolen, J. Hermans, F. Maes, J. Verstraete, F. Van Den Heuvel, T. Depuydt, R. Oyen, and K. Haustermans, *A semi-automated*

2D/3D marker-based registration algorithm modelling prostate shrinkage during radiotherapy for prostate cancer, Radiotherapy and Oncology, 90 (2009), pp. 331–6.

[37] P. J. Burt, *Fast filter transforms for image processing*, Computer Graphics and Image Processing, 16 (1981), pp. 20–51.

[38] T. E. Byrne, *A review of prostate motion with considerations for the treatment of prostate cancer*, Medical Dosimetry, 30 (2005), pp. 155–61.

[39] J. Carl, J. Nielsen, M. Holmberg, E. H. Larsen, K. Fabrin, and R. V. Fisker, *A new fiducial marker for image-guided radiotherapy of prostate cancer: Clinical experience*, Acta Oncologica, 47 (2008), pp. 1358–66.

[40] M. Chamberland, R. Wassenaar, B. Spencer, and T. Xu, *Performance evaluation of real-time motion tracking using positron emission fiducial markers*, Medical Physics, 38 (2011), pp. 810–9.

[41] H. M. Chan, A. C. S. Chung, S. C. H. Yu, and W. M. Wells III, *2D-3D vascular registration between digital subtraction angiographic (DSA) and magnetic resonance angiographic (MRA) images*, in IEEE ISBI: Nano to Macro, vol. 1, 2004, pp. 708–11.

[42] X. Chen, R. C. Gilkeson, and B. Fei, *Automatic 3D-to-2D registration for CT and dual-energy digital radiography for calcification detection*, Medical Physics, 34 (2007), pp. 4934–43.

[43] B. Cho, P. R. Poulsen, A. Sloutsky, A. Sawant, and P. J. Keall, *First demonstration of combined kV/MV image-guided real-time dynamic multileaf-collimator target tracking*, International Journal of Radiation Oncology, Biology, Physics, 74 (2009), pp. 859–67.

[44] B. Cho, Y. Suh, S. Dieterich, and P. J. Keall, *A monoscopic method for real-time tumour tracking using combined occasional x-ray imaging and continuous respiratory monitoring*, Physics in Medicine and Biology, 53 (2008), pp. 2837–55.

[45] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, John Wiley & Sons, Hoboken, NJ, 3 ed., 2008.

[46] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, G. Marchal, K. U. Leuven, and U. Z. Gasthuisberg, *Automated multi-modality image registration based on information theory*, in Information Processing in Medical Imaging, vol. 3, 1995, pp. 263–74.

[47] R. T. Collins, *Mean-shift blob tracking through scale space*, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, IEEE, 2003, pp. II–234.

[48] J. D. Cox, *Fractionation: A paradigm for clinical research in radiation oncology*, International Journal of Radiation Oncology, Biology, Physics, 13 (1987), pp. 1271–81.

[49] Y. Cui, J. G. Dy, B. Alexander, and S. B. Jiang, *Fluoroscopic gating without implanted fiducial markers for lung cancer radiotherapy based on support vector machines*, Physics in Medicine and Biology, 53 (2008), pp. N315–27.

[50] Y. Cui, J. G. Dy, G. C. Sharp, B. Alexander, and S. B. Jiang, *Robust fluoroscopic respiratory gating for lung cancer radiotherapy without implanted fiducial markers*, Physics in Medicine and Biology, 52 (2007), pp. 741–55.

[51] Y. Ding, Y.-C. Chung, S. V. Raman, and O. P. Simonetti, *Application of the Karhunen-Loeve transform temporal image filter to reduce noise in real-time cardiac cine MRI*, Physics in Medicine and Biology, 54 (2009), pp. 3909–22.

[52] S. Eberl, I. Kanno, R. R. Fulton, A. Ryan, B. F. Hutton, and M. J. Fulham, *Automated interstudy image registration technique for SPECT and PET*, The Journal of Nuclear Medicine, 37 (1995), pp. 137–45.

[53] P. M. Evans, *Anatomical imaging for radiotherapy*, Physics in Medicine and Biology, 53 (2008), pp. R151–91.

[54] P. Fallavollita, C. Burdette, D. Song, P. Abolmaesumi, and G. Fichtinger, *C-arm pose estimation in prostate brachytherapy by registration to ultrasound*, in MICCAI, T. Jiang, ed., vol. LNCS 6363, Springer Berlin Heidelberg, 2010, pp. 311–8.

[55] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters, 27 (2006), pp. 861–874.

[56] E. K. Fishman, D. R. Ney, D. G. Heath, F. M. Corl, K. M. Horton, and P. T. Johnson, *Volume rendering versus maximum intensity projection in CT angiography: What works best, when, and why*, RadioGraphics, 26 (2006), pp. 905–23.

[57] J. M. Fitzpatrick, D. L. G. Hill, and C. R. Maurer, *Image registration*, in Handbook of Medical Imaging. Medical Image Processing and Analysis, J. M. Fitzpatrick and M. Sonka, eds., vol. 2, SPIE Press Book, 2000, pp. 449–514.

[58] J. M. Fitzpatrick and J. B. West, *The distribution of target registration error in rigid-body point-based registration*, IEEE Transactions on Medical Imaging, 20 (2001), pp. 917–27.

[59] W. Fledelius, E. Worm, U. V. Elstrøm, J. B. Petersen, C. Grau, M. Høyer, and P. R. Poulsen, *Robust automatic segmentation of multiple implanted cylindrical gold fiducial markers in cone-beam CT projections*, Medical Physics, 38 (2011), pp. 6351–61.

[60] W. Fledelius, E. Worm, M. Høyer, C. Grau, and P. R. Poulsen, *Real-time segmentation of multiple implanted cylindrical liver markers in kilovoltage and megavoltage x-ray images*, Physics in Medicine and Biology, 59 (2014), pp. 2787–800.

[61] E. C. Ford, G. S. Mageras, E. Yorke, K. E. Rosenzweig, R. Wagman, and C. C. Ling, *Evaluation of respiratory movement during gated radiotherapy using film and electronic portal imaging.*, International Journal of Radiation Oncology, Biology, Physics, 52 (2002), pp. 522–31.

[62] B. A. Fraass and A. Eisbruch, *Conformal therapy: Threatment planning, treatment delivery, and clinical results*, in Clinical Radiation Oncology, L. L. Gunderson and J. E. Tepper, eds., Elsevier Churchill Livingston, Philadelphia, PA, 2 ed., 2007.

[63] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, *Multiscale vessel enhancement filtering*, in MICCAI, W. M. Wells, A. Colchester, and S. Delp, eds., vol. LNCS 1496, Springer Berlin Heidelberg, 1998, pp. 130–7.

[64] D. Fu and G. Kuduvalli, *A fast, accurate, and automatic 2D-3D image registration for image-guided cranial radiosurgery*, Medical Physics, 35 (2008), pp. 2180–94.

[65] B. Gagel, C. Demirel, A. Kientopf, M. Pinkawa, M. Piroth, S. Stanzel, C. Breuer, B. Asadpour, T. Jansen, R. Holy, J. E. Wildberger, and M. J. Eble, *Active breathing control (ABC): Determination and reduction of breathing-induced organ motion in the chest*, International Journal of Radiation Oncology, Biology, Physics, 67 (2007), pp. 742–9.

[66] P. R. Geraghty, S. T. Kee, G. McFarlane, M. K. Razavi, D. Y. Sze, and M. D. Dake, *CT-guided transthoracic needle aspiration biopsy of pulmonary nodules: Needle size and pneumothorax rate*, Radiology, 229 (2003), pp. 475–81.

[67] M. Ghilezan, D. Yan, and A. Martinez, *Adaptive radiation therapy for prostate cancer*, Seminars in Radiation Oncology, 20 (2010), pp. 130–7.

[68] M. J. Ghilezan, D. A. Jaffray, J. H. Siewerdsen, M. Van Herk, A. Shetty, M. B. Sharpe, S. Zafar Jafri, F. A. Vicini, R. C. Matter, D. S. Brabbins, and A. A. Martinez, *Prostate gland motion assessed with cine-magnetic resonance imaging (cine-MRI)*, International Journal of Radiation Oncology, Biology, Physics, 62 (2005), pp. 406–17.

[69] A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath, *Brain functional localization: A survey of image registration techniques*, IEEE Transactions on Medical Imaging, 26 (2007), pp. 427–51.

[70] A. Gholipour, N. Kehtarnavaz, S. Yousefi, K. Gopinath, and R. Briggs, *Symmetric deformable image registration via optimization of information theoretic measures*, Image and Vision Computing, 28 (2010), pp. 965–75.

[71] P. Giraud, E. Yorke, S. Jiang, L. Simon, K. Rosenzweig, and G. Mageras, *Reduction of organ motion effects in IMRT and conformal 3D radiation delivery by using gating and tracking techniques*, Cancer/Radiothérapie, 10 (2006), pp. 269–82.

[72] R. H. Gong, A. J. Stewart, and P. Abolmaesumi, *A new method for CT to fluoroscope registration based on unscented Kalman filter*, in MICCAI, R. Larsen,

M. Nielsen, and J. Sporring, eds., vol. LNCS 4190, Springer Berlin Heidelberg, 2006, pp. 891–8.

[73] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Morphological image processing*, in Digital Image Processing using Matlab, Pearson Prentice Hall, Upper Saddle River, NJ, 2004, ch. 9, pp. 334–377.

[74] R. Graf, P. Wust, V. Budach, and D. Boehmer, *Potentials of on-line repositioning based on implanted fiducial markers and electronic portal imaging in prostate cancer radiotherapy*, Radiation Oncology, 4 (2009), pp. 1–9.

[75] J. V. Hajnal, D. J. Hawkes, and D. L. G. Hill, eds., *Medical Image Registration*, CRC Press, 2001.

[76] J. V. Hajnal, N. Saeed, A. Oatridge, E. J. Williams, I. R. Young, and G. M. Bydder, *Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images*, Journal of Computer Assisted Tomography, 19 (1995), pp. 677–91.

[77] E. J. Hall, *The crooked shall be made straight; dose-response relationships for carcinogenesis*, International Journal of Radiation Biology, 80 (2004), pp. 327–37.

[78] N. M. Hamming, M. J. Daly, J. C. Irish, and J. H. Siewerdsen, *Automatic image-to-world registration based on x-ray projections in cone-beam CT-guided interventions*, Medical Physics, 36 (2009), pp. 1800–12.

[79] B. Hannaford and S. A. Glantz, *Adaptive linear predictor tracks implanted radiopaque markers*, IEEE Transactions on Biomedical Engineering, 32 (1985), pp. 117–25.

[80] F. J. Harris, *On the use of windows for harmonic analysis with the discrete Fourier Transform*, Proceedings of the IEEE, 66 (1978), pp. 51–83.

[81] D. J. Hawkes, *Algorithms for radiological image registration and their clinical application*, Journal of Anatomy, 193 (1998), pp. 347–61.

[82] M. A. Hawkins, K. K. Brock, C. Eccles, D. Moseley, D. Jaffray, and L. A. Dawson, *Assessment of residual error in liver position using kV cone-beam computed tomography for liver cancer high-precision radiation therapy*, International Journal of Radiation Oncology, Biology, Physics, 66 (2006), pp. 610–9.

[83] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, *Medical image registration*, Physics in Medicine and Biology, 46 (2001), pp. R1–45.

[84] D. L. G. Hill and D. J. Hawkes, *Across-modality registration using intensity-based cost functions*, in Handbook of Medical Imaging, Processing and Analysis, I. N. Bankman, ed., Academic Press, San Diego, CA, 2000, ch. 34, pp. 537–54.

[85] J. H. Hipwell, G. P. Penney, R. A. Mclaughlin, K. Rhode, P. Summers, T. C. Cox, J. V. Byrne, J. A. Noble, and D. J. Hawkes, *Intensity-based 2-D–3-D registration of cerebral angiograms*, IEEE Transactions on Medical Imaging, 22 (2003), pp. 1417–26.

[86] C. K. Hoh, M. Dahlbom, G. Harris, Y. Choi, R. A. Hawkins, M. E. Phelps, and J. Maddahi, *Automated iterative three-dimensional registration of positron emission tomography images*, Journal of Nuclear Medicine, 34 (1993), pp. 2009–18.

[87] J. D. P. Hoisak, K. E. Sixel, R. Tirona, P. C. F. Cheung, and J.-P. Pignol, *Correlation of lung tumor motion with external surrogate indicators of respiration*, International Journal of Radiation Oncology, Biology, Physics, 60 (2004), pp. 1298–306.

[88] E. Huang, L. Dong, A. Chandra, D. A. Kuban, I. I. Rosen, A. Evans, and A. Pollack, *Intrafraction prostate motion during IMRT for prostate cancer*, International Journal of Radiation Oncology, Biology, Physics, 53 (2002), pp. 261–8.

[89] M. K. Islam, T. G. Purdie, B. D. Norrlinger, H. Alasti, D. J. Moseley, M. B. Sharpe, J. H. Siewerdsen, and D. A. Jaffray, *Patient dose from kilovoltage cone beam computed tomography imaging in radiation therapy*, Medical Physics, 33 (2006), pp. 1573–82.

[90] F. Jacobs, E. Sundermann, B. De Sutter, M. Christiaens, and I. Lemahieu, *A fast algorithm to calculate the exact radiological path through a pixel or voxel space*, Journal of Computing and Information Technology, 6 (1998), pp. 89–94.

[91] D. A. Jaffray, *Emergent technologies for 3-dimensional image-guided radiation delivery*, Seminars in Radiation Oncology, 15 (2005), pp. 208–16.

[92] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, *Flat-panel cone-beam computed tomography for image-guided radiation therapy*, International Journal of Radiation Oncology, Biology, Physics, 53 (2002), pp. 1337–49.

[93] S. B. Jiang, *Technical aspects of image-guided respiration-gated radiation therapy*, Medical Dosimetry, 31 (2006), pp. 141–51.

[94] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Society of Industrial and Applied Mathematics, 2001.

[95] E. Kanoulas, J. A. Aslam, G. C. Sharp, R. I. Berbeco, S. Nishioka, H. Shirato, and S. B. Jiang, *Derivation of the tumor position from external respiratory surrogates with periodical updating of the internal/external correlation*, Physics in Medicine and Biology, 52 (2007), pp. 5443–56.

[96] F. Kanters, T. Denton, A. Shokoufandeh, L. Florack, and B. ter Haar Romeny, *Combining different types of scale space interest points using canonical sets*, in SSVM, LNCS 4485, F. Sgallari, A. Murli, and N. Paragios, eds., Springer Berlin Heidelberg, 2007, pp. 374–85.

[97] I. Kassim, H. Joosten, J. C. Barnhoorn, B. J. M. Heijmen, and M. L. P. Dirkx, *Implications of artefacts reduction in the planning CT originating from implanted fiducial markers*, Medical Dosimetry, 36 (2011), pp. 119–25.

[98] P. J. Keall, A. D. Todor, S. S. Vedam, C. L. Bartee, J. V. Siebers, V. R. Kini, and R. Mohan, *On the use of EPID-based implanted marker tracking for 4D radiotherapy*, Medical Physics, 31 (2004), pp. 3492–9.

[99] F. M. Khan, *Treatment Planning in Radiation Oncology*, Lippincott Williams & Wilkins, Philadelphia, PA, 2 ed., 2007.

[100] D. J. Kim, B. R. Murray, R. Halperin, and W. H. Roa, *Held-breath self-gating technique for radiotherapy of non-small-cell lung cancer: A feasibility study*, International Journal of Radiation Oncology, Biology, Physics, 49 (2001), pp. 43–9.

[101] J. Kim, R. Hammoud, D. Pradhan, H. Zhong, R. Y. Jin, B. Movsas, and I. J. Chetty, *Prostate localization on daily cone-beam computed tomography images: Accuracy assessment of similarity metrics*, International Journal of Radiation Oncology, Biology, Physics, 77 (2010), pp. 1257–65.

[102] J. Kim, F.-F. Yin, Y. Zhao, and J. H. Kim, *Effects of x-ray and CT image enhancements on the robustness and accuracy of a rigid 3D/2D image registration*, Medical Physics, 32 (2005), pp. 866–73.

[103] K. Kitamura, H. Shirato, Y. Seppenwoolde, R. Onimaru, M. Oda, K. Fujita, S. Shimizu, N. Shinohara, T. Harabayashi, and K. Miyasaka, *Three-dimensional intrafractional movement of prostate measured during real-time tumor-tracking radiotherapy in supine and prone treatment positions*, International Journal of Radiation Oncology, Biology, Physics, 53 (2002), pp. 1117–23.

[104] S. Klein, M. Staring, and J. P. W. Pluim, *Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines*, IEEE Transactions on Image Processing, 16 (2007), pp. 2879–90.

[105] M. Koch, J. S. Maltz, S. J. Belongie, B. Gangadharan, S. Bose, H. Shukla, and A. R. Bani-Hashemi, *Automatic coregistration of volumetric images based on implanted fiducial markers*, Medical Physics, 35 (2008), pp. 4513–23.

[106] J. J. Koenderink, *The structure of images*, Biological Cybernetics, 50 (1984), pp. 363–70.

[107] R. Koshani, J. M. Balter, J. A. Hayman, G. T. Henning, and M. van Herk, *Short-term and long-term reproducibility of lung tumor position using active breathing control (ABC)*, International Journal of Radiation Oncology, Biology, Physics, 65 (2006), pp. 1553–9.

[108] A. N. T. J. Kotte, P. Hofman, J. J. W. Lagendijk, M. van Vulpen, and U. A. van der Heide, *Intrafraction motion of the prostate during external-beam radiation therapy: Analysis of 427 patients with implanted fiducial markers*, International Journal of Radiation Oncology, Biology, Physics, 69 (2007), pp. 419–25.

[109] A. Krauss, S. Nill, and U. Oelfke, *The comparative performance of four respiratory motion predictors for real-time tumour tracking*, Physics in Medicine and Biology, 56 (2011), pp. 5303–17.

[110] T. Kron, J. Thomas, C. Fox, A. Thompson, R. Owen, A. Herschtal, A. Haworth, K.-H. Tai, and F. Foroudi, *Intra-fraction prostate displacement in radiotherapy estimated from pre- and post-treatment imaging of patients with implanted fiducial markers*, Radiotherapy and Oncology, 95 (2010), pp. 191–7.

[111] H. D. Kubo and B. C. Hill, *Respiration gated radiotherapy treatment: A technical study*, Physics in Medicine and Biology, 41 (1996), pp. 83–91.

[112] K. S. Kump, G. M. Saidel, and D. L. Wilson, *Comparison of algorithms for combining x-ray angiography images*, IEEE Transactions on Medical Imaging, 20 (2001), pp. 742–50.

[113] P. Kupelian, T. Willoughby, A. Mahadevan, T. Djemil, G. Weinstein, S. Jani, C. Enke, T. Solberg, N. Flores, D. Liu, D. Beyer, and L. Levine, *Multi-institutional clinical experience with the Calypso System in localization and continuous, real-time monitoring of the prostate gland during external radiotherapy*, International Journal of Radiation Oncology, Biology, Physics, 67 (2007), pp. 1088–98.

[114] P. A. Kupelian, T. R. Willoughby, S. L. Meeks, A. Forbes, T. Wagner, M. Maach, and K. M. Langen, *Intraprostatic fiducials for localization of the prostate gland: Monitoring intermarker distances during radiation therapy to test for marker stability*, International Journal of Radiation Oncology, Biology, Physics, 62 (2005), pp. 1291–6.

[115] P. Lacroute and M. Levoy, *Fast volume rendering using a shear-warp factorization of the viewing transformation*, in SIGGRAPH, Orlando, FL, 1994, pp. 451–8.

[116] K. M. Langen and D. T. L. Jones, *Organ motion and its management*, International Journal of Radiation Oncology, Biology, Physics, 50 (2001), pp. 265–78.

[117] K. M. Langen, T. R. Willoughby, S. L. Meeks, A. Santhanam, A. Cunningham, L. Levine, and P. A. Kupelian, *Observations on real-time prostate gland motion using electomagnetic tracking*, International Journal of Radiation Oncology, Biology, Physics, 71 (2008), pp. 1084–90.

[118] H. Lee, J. Lee, N. Kim, I. K. Lyoo, and Y. G. Shin, *Robust and fast shell registration in PET and MR/CT brain images*, Computers in Biology and Medicine, 39 (2009), pp. 961–77.

[119] H. Lester and S. R. Arridge, *A survey of hierarchical non-linear medical image registration*, Pattern Recognition, 32 (1999), pp. 129–49.

[120] M. Levoy, *Efficient ray tracing of volume data*, ACM Transactions on Graphics, 9 (1990), pp. 245–61.

[121] Q. Li, S. Sone, and K. Doi, *Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans*, Medical Physics, 30 (2003), pp. 2040–51.

[122] R. Li and G. Sharp, *Robust fluoroscopic tracking of fiducial markers: Exploiting the spatial constraints*, Physics in Medicine and Biology, 58 (2013), pp. 1789–808.

[123] T. Lin, L. I. Cerviño, X. Tang, N. Vasconcelos, and S. B. Jiang, *Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy*, Physics in Medicine and Biology, 54 (2009), pp. 981–92.

[124] T. Lin, R. Li, X. Tang, J. G. Dy, and S. B. Jiang, *Markerless gating for lung cancer radiotherapy based on machine learning techniques*, Physics in Medicine and Biology, 54 (2009), pp. 1555–63.

[125] T. Lindeberg, *Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention*, International Journal of Computer Vision, 11 (1993), pp. 283–318.

[126] ——, *On scale selection for differential operators*, in 8th Scandinavian Conference on Image Analysis, Tromsø, Norway, 1993, pp. 857–66.

[127] ——, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Dordrecht, 1994.

[128] N. Linthout, S. Bral, I. Van de Vondel, D. Verellen, K. Tournel, T. Gevaert, M. Duchateau, T. Reynders, and G. Storme, *Treatment delivery time optimization of respiratory gated radiation therapy by application of audio-visual feedback*, Radiotherapy and Oncology, 91 (2009), pp. 330–5.

[129] D. W. Litzenberg, J. M. Balter, S. W. Hadley, H. M. Sandler, T. R. Willoughby, P. A. Kupelian, and L. Levine, *Influence of intrafraction motion on margins for prostate radiotherapy*, International Journal of Radiation Oncology, Biology, Physics, 65 (2006), pp. 548–53.

[130] L. Liu, B. Sun, and N. Wei, *A novel marker tracking method based on extended Kalman filter for multi-camera optical tracking systems*, in iCBBE, IEEE, 2011, pp. 1–5.

[131] Á. Logadóttir, S. Korreman, and P. M. Petersen, *Comparison of the accuracy and precision of prostate localization with 2D-2D and 3D images*, Radiotherapy and Oncology, 98 (2011), pp. 175–80.

[132] C. Lorenz, I.-C. Carlsen, T. M. Buzug, C. Fassnacht, and J. Weese, *Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images*, in CRVMed-MRCAS, Springer Berlin Heidelberg, 1997, pp. 233–42.

[133] D. A. Low, M. Nystrom, E. Kalinin, P. Parikh, J. F. Dempsey, J. D. Bradley, S. Mutic, S. H. Wahab, T. Islam, G. Christensen, D. G. Politte, and B. R. Whiting, *A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing*, Medical Physics, 30 (2003), pp. 1254–63.

[134] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60 (2004), pp. 91–110.

[135] F. Maes, D. Vandermeulen, and P. Suetens, *Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information*, Medical Image Analysis, 3 (1999), pp. 373–86.

[136] D. Mah, J. Hanley, K. E. Rosenzweig, E. Yorke, L. Braban, C. C. Ling, S. A. Leibel, and G. Mageras, *Technical aspects of the deep inspiration breath-hold technique in the treatment of thoracic cancer*, International Journal of Radiation Oncology, Biology, Physics, 48 (2000), pp. 1175–85.

[137] A. H. Mahnken, R. Raupach, J. E. Wildberger, B. Jung, N. Heussen, T. G. Flohr, R. W. Günther, and S. Schaller, *A new algorithm for metal artifact reduction in computed tomography – In vitro and in vivo evaluation after total hip replacement*, Investigative Radiology, 38 (2003), pp. 769–75.

[138] J. B. A. Maintz and M. A. Viergever, *A survey of medical image registration*, Medical Image Analysis, 2 (1998), pp. 1–36.

[139] C. D. Manning, P. Raghavan, and H. Schütze, *Hierarchical clustering*, in Introduction to Information Retrieval, Cambridge University Press, New York, 2008, ch. 17, pp. 346–68.

[140] W. Mao, N. Riaz, L. Lee, R. Wiersma, and L. Xing, *A fiducial detection algorithm for real-time image guided IMRT based on simultaneous MV and kV imaging*, Medical Physics, 35 (2008), pp. 3554–64.

[141] W. Mao, R. D. Wiersma, and L. Xing, *Fast internal marker tracking algorithm for onboard MV and kV imaging systems*, Medical Physics, 35 (2008), pp. 1942–9.

[142] T. E. Marchant, A. Skalski, and B. J. Matuszewski, *Automatic tracking of implanted fiducial markers in cone beam CT projection images*, Medical Physics, 39 (2012), pp. 1322–34.

[143] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, *A review of 3D/2D registration methods for image-guided interventions*, Medical Image Analysis, 16 (2012), pp. 642–61.

[144] A. A. Martinez, D. Yan, D. Lockman, D. Brabbins, K. Kota, M. Sharpe, D. A. Jaffray, F. Vicini, and J. Wong, *Improvement in dose escalation using the process of adaptive radiotherapy combined with three-dimensional conformal or intensity-modulated beams for prostate cancer*, International Journal of Radiation Oncology, Biology, Physics, 50 (2001), pp. 1226–34.

[145] C. R. Maurer and J. M. Fitzpatrick, *A review of medical image registration*, in Interactive Image-Guided Neurosurgery, R. J. Maciunas, ed., American Association of Neurological Surgeons, Park Ridge, IL, 1993, pp. 17–44.

[146] A. L. McKenzie, M. van Herk, and B. Mijnheer, *The width of margins in radiotherapy treatment plans*, Physics in Medicine and Biology, 45 (2000), pp. 3331–42.

[147] G. J. Meijer, J. de Klerk, K. Bzdusek, H. A. van den Berg, R. Janssen, M. R. Kaus, P. Rodrigus, and P.-P. van der Toorn, *What CTV-to-PTV margins should be applied for prostate irradiation? Four-dimensional quantitative assessment using model-based deformable image registration techniques*, International Journal of Radiation Oncology, Biology, Physics, 72 (2008), pp. 1416–25.

[148] K. Mikolajczyk and C. Schmid, *Scale & affine invariant interest point detectors*, International Journal of Computer Vision, 60 (2004), pp. 63–86.

[149] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, New York, 2004.

[150] ——, *FAIR: Flexible Algorithms for Image Registration*, SIAM, Philadelphia, PA, 2009.

[151] D. J. Moseley, J. H. Siewerdsen, and D. A. Jaffray, *High-contrast object localization and removal in cone-beam CT*, in Medical Imaging, M. J. Flynn, ed., Bellingham, WA, 2005, SPIE, pp. 40–50.

[152] D. J. Moseley, E. A. White, K. L. Wiltshire, T. Rosewall, M. B. Sharpe, J. H. Siewerdsen, J.-P. Bissonnette, M. Gospodarowicz, P. Warde, C. N. Catton, and D. A. Jaffray, *Comparison of localization performance with implanted fiducial markers and cone-beam computed tomography for on-line image-guided radiotherapy of the prostate*, International Journal of Radiation Oncology, Biology, Physics, 67 (2007), pp. 942–53.

[153] K. Mueller, T. Möller, J. E. I. Swan, R. Crawfis, N. Shareef, and R. Yagel, *Splatting errors and antialiasing*, IEEE Transactions on Visualization and Computer Graphics, 4 (1998), pp. 178–91.

[154] R. Munbodh, D. A. Jaffray, D. J. Moseley, Z. Chen, J. P. S. Knisely, P. Cathier, and J. S. Duncan, *Automated 2D-3D registration of a radiograph and a cone beam CT using line-segment enhancement*, Medical Physics, 33 (2006), pp. 1398–1411.

[155] R. Munbodh, H. D. Tagare, Z. Chen, D. A. Jaffray, D. J. Moseley, J. P. S. Knisely, and J. S. Duncan, *2D-3D registration for prostate radiation therapy based on a statistical model of transmission images*, Medical Physics, 36 (2009), pp. 4555–68.

[156] A. J. Mundt and J. Roeske, *Intensity Modulated Radiation Therapy: A Clinical Perspective*, Decker, Hamilton, ON, 2005.

[157] M. J. Murphy, *Tracking moving organs in real time*, Seminars in Radiation Oncology, 14 (2004), pp. 91–100.

[158] M. J. Murphy, J. Balter, S. Balter, J. A. BenComo, I. J. Das, S. B. Jiang, C.-M. Ma, G. H. Olivera, R. F. Rodebaugh, K. J. Ruchala, H. Shirato, and F.-F. Yin, *The management of imaging dose during image-guided radiotherapy: Report of the AAPM Task Group 75*, Medical Physics, 34 (2007), pp. 4041–63.

[159] T. F. Mutanga, H. C. J. de Boer, G. J. van der Wielen, D. Wentzler, J. Barnhoorn, L. Incrocci, and B. J. M. Heijmen, *Stereographic targeting in prostate radiotherapy: Speed and precision by daily automatic positioning corrections using kilovoltage/megavoltage image pairs*, International Journal of Radiation Oncology, Biology, Physics, 71 (2008), pp. 1074–83.

[160] A. Nederveen, J. Lagendijk, and P. Hofman, *Detection of fiducial gold markers for automatic on-line megavoltage position verification using a marker extraction kernel (MEK)*, International Journal of Radiation Oncology, Biology, Physics, 47 (2000), pp. 1435–42.

[161] A. J. Nederveen, H. Dehnad, U. A. van Der Heide, R. J. A. van Moorselaar, P. Hofman, and J. J. W. Lagendijk, *Comparison of megavoltage position verification for prostate irradiation based on bony anatomy and implanted fiducials*, Radiotherapy and Oncology, 68 (2003), pp. 81–8.

[162] A. J. Nederveen, J. J. W. Lagendijk, and P. Hofman, *Feasibility of automatic marker detection with an a-Si flat-panel imager*, Physics in Medicine and Biology, 46 (2001), pp. 1219–30.

[163] A. J. Nederveen, U. A. van Der Heide, H. Dehnad, R. J. A. van Moorselaar, P. Hofman, and J. J. W. Lagendijk, *Measurements and clinical consequences of prostate motion during a radiotherapy fraction*, International Journal of Radiation Oncology, Biology, Physics, 53 (2002), pp. 206–14.

[164] J. A. Ng, J. T. Booth, P. R. Poulsen, W. Fledelius, E. S. Worm, T. Eade, F. Hegi, A. Kneebone, Z. Kuncic, and P. J. Keall, *Kilovoltage intrafraction monitoring for prostate intensity modulated arc therapy: First clinical results*, International Journal of Radiation Oncology, Biology, Physics, 84 (2012), pp. e655–61.

[165] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[166] R. Onimaru, H. Shirato, H. Aoyama, K. Kitakura, T. Seki, K. Hida, K. Fujita, K. Kagei, T. Nishioka, T. Kunieda, Y. Iwasaki, and K. Miyasaka, *Calculation of rotational setup error using the real-time tracking radiation therapy (RTRT) system and its application to the treatment of spinal schwannoma*, International Journal of Radiation Oncology, Biology, Physics, 54 (2002), pp. 939–47.

[167] C. Palm, A. Vieten, D. Salber, and U. Pietrzyk, *Evaluation of registration strategies for multi-modality images of rat brain slices*, Physics in Medicine and Biology, 54 (2009), pp. 3269–89.

[168] R. B. Paranjape, *Fundamental enhancement techniques*, in Handbook of Medical Imaging, Processing and Analysis, I. N. Bankman, ed., Academic Press, San Diego, CA, 2000, ch. 1, pp. 3–19.

[169] J. C. Park, S. H. Park, J. H. Kim, S. M. Yoon, S. S. Kim, J. S. Kim, Z. Liu, T. Watkins, and W. Y. Song, *Four-dimensional cone-beam computed tomography and digital tomosynthesis reconstructions using respiratory signals extracted from transcutaneously inserted metal markers for liver SBRT*, Medical Physics, 38 (2011), pp. 1028–36.

[170] S.-J. Park, D. Ionascu, F. Hacker, H. Mamon, and R. Berbeco, *Automatic marker detection and 3D position reconstruction using cine EPID images for SBRT verification*, Medical Physics, 36 (2009), pp. 4536–46.

[171] G. P. Penney, J. A. Schnabel, D. Rueckert, M. A. Viergever, and W. J. Niessen, *Registration-based interpolation*, IEEE Transactions on Medical Imaging, 23 (2004), pp. 922–6.

[172] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes, *A comparison of similarity measures for use in 2-D–3-D medical image registration*, IEEE Transactions on Medical Imaging, 17 (1998), pp. 586–95.

[173] O. Peshko, T. N. Davidson, J. Modersitzki, T. Terlaky, and D. J. Moseley, *A novel marker enhancement filter (MEF) for fluoroscopic images*, in Journal of Physics: Conference Series, vol. 489, IOP Publishing, 2014, p. 012038.

[174] O. Peshko, D. J. Moseley, T. Craig, T. Terlaky, and C. Menard, *Intra-fraction prostate motion evaluation with on-line fluoroscopic imaging*, Medical Physics, 34 (2007), p. 2392.

[175] J. D. Pintér, *Global Optimization in Action*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[176] M. M. Poggi, D. A. Gant, W. Sewchand, and W. B. Warlick, *Marker seed migration in prostate localization*, International Journal of Radiation Oncology, Biology, Physics, 56 (2003), pp. 1248–51.

[177] P. R. Poulsen, B. Cho, D. Ruan, A. Sawant, and P. J. Keall, *Dynamic multileaf collimator tracking of respiratory target motion based on a single kilovoltage imager during arc radiotherapy*, International Journal of Radiation Oncology, Biology, Physics, 77 (2010), pp. 600–7.

[178] P. R. Poulsen, W. Fledelius, P. J. Keall, E. Weiss, J. Lu, E. Brackbill, and G. D. Hugo, *A method for robust segmentation of arbitrarily shaped radiopaque structures in cone-beam CT projections*, Medical Physics, 38 (2011), pp. 2151–6.

[179] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, 3 ed., 2007.

[180] J. Rogowska, *Overview and fundamentals of medical image segmentation*, in Handbook of Medical Imaging, Processing and Analysis, I. N. Bankman, ed., Academic Press, San Diego, CA, 2000, ch. 5, pp. 69–86.

[181] T. Rohlfing, D. B. Russakoff, J. Denzler, K. Mori, and C. R. Maurer, *Progressive attenuation fields: Fast 2D-3D image registration without precomputation*, Medical Physics, 32 (2005), pp. 2870–80.

[182] T. Rosewall, P. Chung, A. Bayley, G. Lockwood, H. Alasti, R. Bristow, V. Kong, M. Milosevic, and C. Catton, *A randomized comparison of interfraction and intrafraction prostate motion with and without abdominal compression*, Radiotherapy and Oncology, 88 (2008), pp. 88–94.

[183] J. Rottmann, P. Keall, and R. Berbeco, *Markerless EPID image guided dynamic multi-leaf collimator tracking for lung tumors*, Physics in Medicine and Biology, 58 (2013), pp. 4195–204.

[184] D. Ruijters, B. M. ter Haar Romeny, and P. Suetens, *Vesselness-based 2D–3D registration of the coronary arteries*, International Journal of Computer Assisted Radiology and Surgery, 4 (2009), pp. 391–7.

[185] D. B. Russakoff, T. Rohlfing, D. Rueckert, R. Shahidi, D. Kim, and C. R. Maurer, *Fast calculation of digitally reconstructed radiographs using light fields*, in Medical Imaging, M. Sonka and J. M. Fitzpatrick, eds., vol. 5032, SPIE, 2003, pp. 684–95.

[186] Y. Salih and A. S. Malik, *Comparison of stochastic filtering methods for 3D tracking*, Pattern Recognition, 44 (2011), pp. 2711–37.

[187] D. Sarrut, *Geometrical transformation approximation for 2D/3D intensity-based registration of portal images and CT scan*, in MICCAI, W. Niessen and M. Viergever, eds., vol. LNCS 2208, Springer Berlin Heidelberg, 2001, pp. 532–40.

[188] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, *Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images*, Medical Image Analysis, 2 (1998), pp. 143–68.

[189] Y. Sato, C.-F. Westin, A. Bhalerao, S. Nakajima, N. Shiraga, S. Tamura, and R. Kikinis, *Tissue classification based on 3D local intensity structures for volume rendering*, IEEE Transactions on Visualization and Computer Graphics, 6 (2000), pp. 160–80.

[190] J. M. Schallenkamp, M. G. Herman, J. J. Kruse, and T. M. Pisansky, *Prostate position relative to pelvic bony anatomy based on intraprostatic gold markers and electronic portal imaging*, International Journal of Radiation Oncology, Biology, Physics, 63 (2005), pp. 800–11.

[191] B. Schaly, G. S. Bauman, W. Song, J. J. Battista, and J. Van Dyk, *Dosimetric impact of image-guided 3D conformal radiation therapy of prostate cancer*, Physics in Medicine and Biology, 50 (2005), pp. 3083–101.

[192] G. Schoonenberg, M. Schrijver, Q. Duan, R. Kemkers, and A. Laine, *Adaptive spatial-temporal filtering applied to x-ray fluoroscopy angiography*, in Medical Imaging, R. L. Galloway and K. R. Cleary, eds., vol. 5744, Bellingham, WA, 2005, SPIE, pp. 870–8.

[193] A. Schweikard, G. Glosser, M. Bodduluri, M. J. Murphy, and J. R. Adler, *Robotic motion compensation for respiratory movement during radiosurgery*, Computer Aided Surgery, 5 (2000), pp. 263–77.

[194] A. Schweikard, H. Shiomi, and J. Adler, *Respiration tracking in radiosurgery*, Medical Physics, 31 (2004), pp. 2738–41.

[195] G. C. Sharp, S. B. Jiang, S. Shimizu, and H. Shirato, *Prediction of respiratory tumour motion for real-time image-guided radiotherapy*, Physics in Medicine and Biology, 49 (2004), pp. 425–40.

[196] ——, *Tracking errors in a prototype real-time tumour tracking system*, Physics in Medicine and Biology, 49 (2004), pp. 5347–56.

[197] A. Shen and L. Luo, *Point-based digitally reconstructed radiograph*, in ICPR, IEEE, 2008.

[198] H. Shirato, S. Shimizu, K. Kitamura, T. Nishioka, K. Kagei, S. Hashimoto, H. Aoyama, T. Kunieda, N. Shinohara, H. Dosaka-Akita, and K. Miyasaka, *Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor*, International Journal of Radiation Oncology, Biology, Physics, 48 (2000), pp. 435–42.

[199] H. Shirato, S. Shimizu, T. Kunieda, K. Kitamura, M. van Herk, K. Kagei, T. Nishioka, S. Hashimoto, K. Fujita, H. Aoyama, K. Tsuchiya, K. Kudo, and K. Miyasaka, *Physical aspects of a real-time tumor-tracking system for gated radiotherapy*, International Journal of Radiation Oncology, Biology, Physics, 48 (2000), pp. 1187–95.

[200] S. Siddique, E. Fiume, and D. A. Jaffray, *Minimizing dose during fluoroscopic tracking through geometric performance feedback*, Medical Physics, 38 (2011), pp. 2494–507.

[201] R. L. Siddon, *Fast calculation of the exact radiological path for a three-dimensional CT array*, Medical Physics, 12 (1985), pp. 252–5.

[202] J. H. Siewerdsen, D. J. Moseley, B. Bakhtiar, S. Richard, and D. A. Jaffray, *The influence of antiscatter grids on soft-tissue detectability in cone-beam computed tomography with flat-panel detectors*, Medical Physics, 31 (2004), pp. 3506–20.

[203] D. Škerl, B. Likar, and F. Pernuš, *A protocol for evaluation of similarity measures for rigid registration*, IEEE Transactions on Medical Imaging, 25 (2006), pp. 779–91.

[204] P. Slagmolen, J. Hermans, F. Maes, T. Budiharto, K. Haustermans, and F. van den Heuvel, *Fast, accurate, and robust automatic marker detection for motion correction based on oblique kV or MV projection image pairs*, Medical Physics, 37 (2010), pp. 1554–64.

[205] M. H. P. Smitsmans, J. de Bois, J.-J. Sonke, A. Betgen, L. J. Zijp, D. A. Jaffray, J. V. Lebesque, and M. van Herk, *Automatic prostate localization on cone-beam CT scans for high precision image-guided radiotherapy*, International Journal of Radiation Oncology, Biology, Physics, 63 (2005), pp. 975–84.

[206] M. H. P. Smitsmans, F. J. Pos, J. de Bois, W. D. Heemsbergen, J.-J. Sonke, J. V. Lebesque, and M. van Herk, *The influence of a dietary protocol on cone beam CT-guided radiotherapy for prostate cancer patients*, International Journal of Radiation Oncology, Biology, Physics, 71 (2008), pp. 1279–86.

[207] J.-J. Sonke, L. Zijp, P. Remeijer, and M. van Herk, *Respiratory correlated cone beam CT*, Medical Physics, 32 (2005), pp. 1176–86.

[208] A. Søvik, E. Malinen, and D. R. Olsen, *Strategies for biologic image-guided dose escalation: A review*, International Journal of Radiation Oncology, Biology, Physics, 73 (2009), pp. 650–8.

[209] G. Strang and T. Nguyen, *Wavelets and filter banks*, SIAM, 1996.

[210] J. S. Stromberg, M. B. Sharpe, L. H. Kim, V. R. Kini, D. A. Jaffray, A. A. Martinez, and J. W. Wong, *Active breathing control (ABC) for Hodgkin's disease: Reduction in normal tissue irradiation with deep inspiration and implications for treatment*, International Journal of Radiation Oncology, Biology, Physics, 48 (2000), pp. 797–806.

[211] C. Studholme, D. L. G. Hill, and D. J. Hawkes, *An overlap invariant entropy measure of 3D medical image alignment*, Pattern Recognition, 32 (1999), pp. 71–86.

[212] W. Sun, W. J. Niessen, and S. Klein, *Hierarchical vs. simultaneous multiresolution strategies*, in WBIR, B. M. Dawant, ed., vol. LNCS 7359, Springer Berlin Heidelberg, 2012, pp. 60–9.

[213] T. S. Y. Tang, N. J. MacIntyre, H. S. Gill, R. A. Fellows, N. A. Hill, D. R. Wilson, and R. E. Ellis, *Accurate assessment of patellar tracking using fiducial and intensity-based fluoroscopic techniques*, Medical Image Analysis, 8 (2004), pp. 343–51.

[214] X. Tang, G. C. Sharp, and S. B. Jiang, *Fluoroscopic tracking of multiple implanted fiducial markers using multiple object tracking*, Physics in Medicine and Biology, 52 (2007), pp. 4081–98.

[215] T. Tuytelaars and K. Mikolajczyk, *Local invariant feature detectors: A survey*, Foundations and Trends in Computer Graphics and Vision, 3 (2007), pp. 177–280.

[216] A. Vaccarella, E. De Momi, A. Enquobahrie, and G. Ferrigno, *Unscented Kalman filter based sensor fusion for robust optical and electromagnetic tracking in surgical navigation*, IEEE Transactions on Instrumentation and Measurement, 62 (2013), pp. 2067–81.

[217] U. A. van der Heide, A. N. T. J. Kotte, H. Dehnad, P. Hofman, J. J. W. Lagenijk, and M. van Vulpen, *Analysis of fiducial marker-based position verification in the external beam radiotherapy of patients with prostate cancer*, Radiotherapy and Oncology, 82 (2007), pp. 38–45.

[218] P. M. A. van Haaren, A. Bel, P. Hofman, M. van Vulpen, A. N. T. J. Kotte, and U. A. van der Heide, *Influence of daily setup measurements and corrections on the estimated delivered dose during IMRT treatment of prostate cancer patients*, Radiotherapy and Oncology, 90 (2009), pp. 291–8.

[219] M. van Herk, *Errors and margins in radiotherapy*, Seminars in Radiation Oncology, 14 (2004), pp. 52–64.

[220] M. van Herk, P. Remeijer, C. Rasch, and J. V. Lebesque, *The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy*, International Journal of Radiation Oncology, Biology, Physics, 47 (2000), pp. 1121–35.

[221] S. S. Vedam, P. J. Keall, A. Docef, D. a. Todor, V. R. Kini, and R. Mohan, *Predicting respiratory motion for four-dimensional radiotherapy*, Medical Physics, 31 (2004), pp. 2274–83.

[222] S. S. Vedam, V. R. Kini, P. J. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan, *Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker*, Medical Physics, 30 (2003), pp. 505–13.

[223] P. S. Verma, H. Wu, M. P. Langer, I. J. Das, and G. Sandison, *Survey: Real-time tumor motion prediction for image-guided radiation treatment*, Computing in Science and Engineering, 13 (2011), pp. 24–35.

[224] E. Vigneault, J. Pouliot, J. Laverdiere, J. Roy, and M. Dorion, *Electronic portal imaging device detection of radioopaque markers for the evaluation of prostate position during megavoltage irradiation: A clinical study*, International Journal of Radiation Oncology, Biology, Physics, 37 (1997), pp. 205–12.

[225] P. A. Viola, *Alignment by Maximization of Mutual Information*, PhD thesis, Massachusetts Institute of Technology, 1995.

[226] R. Wagman, E. Yorke, E. Ford, P. Giraud, G. Mageras, B. Minsky, and K. Rosenzweig, *Respiratory gating for liver tumors: Use in dose escalation*, International Journal of Radiation Oncology, Biology, Physics, 55 (2003), pp. 659–68.

[227] J. Wang, L. Zhu, and L. Xing, *Noise reduction in low-dose x-ray fluoroscopy for image-guided radiation therapy*, International Journal of Radiation Oncology, Biology, Physics, 74 (2009), pp. 637–43.

[228] M. Y. Wang, J. M. Fitzpatrick, and C. R. Maurer, *Design of fiducials for accurate registration of CT and MR volume images*, in Medical Imaging, vol. 2434, SPIE, 1995, pp. 96–108.

[229] Z. Wang, C. G. Willett, and F.-F. Yin, *Reduction of organ motion by combined cardiac gating and respiratory gating*, International Journal of Radiation Oncology, Biology, Physics, 68 (2007), pp. 259–66.

[230] S. Webb, *Motion effects in (intensity modulated) radiation therapy: A review*, Physics in Medicine and Biology, 51 (2006), pp. R403–25.

[231] J. Weese, T. M. Buzug, G. P. Penney, and P. Desmedt, *2D/3D registration and motion tracking for surgical interventions*, Philips Journal of Research, 51 (1998), pp. 299–316.

[232] L. Westover, *Footprint evaluation for volume rendering*, Computer Graphics, 24 (1990), pp. 367–76.

[233] R. D. Wiersma, W. Mao, and L. Xing, *Combined kV and MV imaging for real-time tracking of implanted fiducial markers*, Medical Physics, 35 (2008), pp. 1191–8.

[234] R. D. Wiersma, N. Riaz, S. Dieterich, Y. Suh, and L. Xing, *Use of MV and kV imager correlation for maintaining continuous real-time 3D internal marker tracking during beam interruptions*, Physics in Medicine and Biology, 54 (2009), pp. 89–103.

[235] A. P. Witkin, *Scale-space filtering: A new approach to multi-scale description*, in IEEE ICASSP, vol. 9, IEEE, 1984, pp. 150–3.

[236] J. W. H. Wolthaus, M. van Herk, S. H. Muller, J. S. a. Belderbos, J. V. Lebesque, J. A. de Bois, M. M. G. Rossi, and E. M. F. Damen, *Fusion of respiration-correlated PET and CT scans: Correlated lung tumour motion in anatomical and functional scans*, Physics in Medicine and Biology, 50 (2005), pp. 1569–83.

[237] J. W. Wong, M. B. Sharpe, D. A. Jaffray, V. R. Kini, J. M. Robertson, J. S. Stromberg, and A. A. Martinez, *The use of active breathing control (ABC) to reduce margin for breathing motion*, International Journal of Radiation Oncology, Biology, Physics, 44 (1999), pp. 911–9.

[238] R. P. Woods, *Spatial transformation models*, in Handbook of Medical Imaging, Processing and Analysis, I. N. Bankman, ed., Academic Press, San Diego, CA, 2000, ch. 29, pp. 465–90.

[239] ——, *Within-modality registration using intensity-based cost functions*, in Handbook of Medical Imaging, Processing and Analysis, I. N. Bankman, ed., Academic Press, San Diego, CA, 2000, ch. 33, pp. 529–36.

[240] H. Wu, Q. Zhao, R. I. Berbeco, S. Nishioka, H. Shirato, and S. B. Jiang, *Gating based on internal/external signals with dynamic correlation updates*, Physics in Medicine and Biology, 53 (2008), pp. 7137–50.

[241] J. Wu, M. Kim, J. Peters, H. Chung, and S. S. Samant, *Evaluation of similarity measures for use in the intensity-based rigid 2D-3D registration for patient positioning in radiotherapy*, Medical Physics, 36 (2009), pp. 5391–403.

[242] W. Wunderink, A. Méndez Romero, W. de Kruijf, H. de Boer, P. Levendag, and B. Heijmen, *Reduction of respiratory liver tumor motion by abdominal compression in stereotactic body frame, analyzed by tracking fiducial markers implanted in liver*, International Journal of Radiation Oncology, Biology, Physics, 71 (2008), pp. 907–15.

[243] L. Xing, B. Thorndyke, E. Schreibmann, Y. Yang, T.-F. Li, G.-Y. Kim, G. Luxton, and A. Koong, *Overview of image-guided radiation therapy*, Medical Dosimetry, 31 (2006), pp. 91–112.

[244] T. Xu, J. T. Wong, P. M. Shikhaliev, J. L. Ducote, M. S. Al-Ghazi, and S. Molloi, *Real-time tumor tracking using implanted positron emission markers: Concept and simulation study*, Medical Physics, 33 (2006), pp. 2598–609.

[245] M. Yazdia, L. Gingras, and L. Beaulieu, *An adaptive approach to metal artifact reduction in helical computer tomography for radiation therapy treatment planning: Experimental and clinical results*, International Journal of Radiation Oncology, Biology, Physics, 62 (2005), pp. 1224–31.

[246] E. M. Zeman, *Biological basis of radiation oncology*, in Clinical Radiation Oncology, L. L. Gunderson and J. E. Tepper, eds., Elsevier Churchill Livingston, Philadelphia, PA, 2 ed., 2007.

[247] T. Zhang, H. Keller, M. J. O'Brien, T. R. Mackie, and B. Paliwal, *Application of the spirometer in respiratory gated radiotherapy*, Medical Physics, 30 (2003), pp. 3165–71.

[248] Y. Zhang, L. Zhang, X. R. Zhu, A. K. Lee, M. Chambers, and L. Dong, *Reducing metal artifacts in cone-beam CT images by preprocessing projection data*, International Journal of Radiation Oncology, Biology, Physics, 67 (2007), pp. 924–32.

[249] L. Zollei, *2D-3D Rigid-Body Registration of X-Ray Fluoroscopy and CT Images*, master thesis, Massachusetts Institute of Technology, 2001.

[250] M. Zwicker, H. Pfister, J. van Baar, and M. Gross, *EWA splatting*, IEEE Transactions on Visualization and Computer Graphics, 8 (2002), pp. 223–38.

Appendix A

System Hardware and Software

The computing system used to conduct experiments in this thesis possessed the following characteristics:

Processor:	Intel® Core™2 Quad CPU Q6700 @ 2.66GHz
RAM:	4 Gb
Operating system:	Windows 7 Professional, 64-bit
Matlab:	R2009b (7.9.0.529) 32-bit

The use of this particular Matlab version is due to the necessity of running precompiled libraries for opening the image files of the XVI format (CBCT images with all supporting information). To be able to access information on patients, treatments, and images in the testing data sets from Matlab, we were bound to its 32-bit version, as the Microsoft dBase driver was not available in the 64-bit version.

Appendix B

Geometric Transformations

This appendix provides details on geometric transformations and coordinate conventions used in this thesis. Reference [238] contains an extensive discussion of coordinate transformations with formulations and examples.

One of the ways to express a *rigid* geometric transformation that acts on a point $p \in \mathbb{R}^d$ to produce a new point $q \in \mathbb{R}^d$ is through the $d \times d$ matrix of *rotation* R and the d-element vector of *translation* (shift) t as $q = Rp + t$, or $q = R(p + t)$ if the translation has to be applied first. We find it more convenient, however, to be able to have a unified representation of the transformations as matrices of the same size, especially taking into account that in addition to translations and rotations we also use a perspective transformation that is involved in 2D-3D image registration. Such representation is made possible through the use of *homogeneous coordinates* [238]. In homogeneous coordinates, points in $\mathbb{R}^d$ are expressed as vectors in $\mathbb{R}^{d+1}$, e.g., a point $p = (p_x, p_y, p_z)^T$ can be interpreted in homogeneous coordinates as $\tilde{p} = (p_x, p_y, p_z, 1)^T$. Correspondingly, all transformations are expressed as $\mathbb{R}^{(d+1)\times(d+1)}$ matrices [238, 57], where the detailed definitions for our transformations are given later in this appendix. After the transformations are applied, the point $\tilde{q} = (q_x, q_y, q_z, q_w)^T$ can be converted back to $\mathbb{R}^3$ by dividing by its last coordinate $q_w \neq 0$ to obtain $q = (q_x/q_w, q_y/q_w, q_z/q_w)^T$. The homogeneous coordinate transformation for $p \in \mathbb{R}^2$ is analogous. In this thesis, we may implicitly switch to homogeneous coordinates before applying geometric transformations, and back after. However, in this appendix, the notation with the tilde will be explicitly used to denote the points in homogeneous coordinates.

B.1 3D Rigid Geometric Transformation in XYZ System

As explained in Section 2.4.2, XYZ is a 3D *right-handed* coordinate system fixed in the room with the origin at the LINAC's centre of rotation, the *isocentre*. The coordinates of the grid points of the 3D CBCT images are recorded in the XYZ system by the XVI software during volumetric image reconstruction from the 2D x-ray projections. As explained in the problem setup in Section 2.2, the motion of a target is modelled by the rigid transformations, i.e., by rotations and translations. A transformed point $q \in \mathbb{R}^3$ is obtained from $p \in \mathbb{R}^3$ as:

$$q = \mathcal{T}^{3D}(a, p), \text{ or } \tilde{q} = R^{3D}T^{3D}\tilde{p}, \tag{B.1.1}$$

where $\tilde{p}, \tilde{q} \in \mathbb{R}^4$ are the points p and q in homogeneous coordinates, $\mathcal{T}^{3D} : \mathbb{R}^6 \times \mathbb{R}^3 \to \mathbb{R}^3$ is the 3D rigid transformation with a vector of parameters $a = (t_x, t_y, t_z, \gamma_x, \gamma_y, \gamma_z)^T$ (three shifts and three rotation angles), and R^{3D} and T^{3D} are the 4×4 matrices of 3D rotation and translation, respectively, defined as explained below.

According to the Euler's rotation theorem, any rotation can be described using three angles, which means that decomposition into three matrices can be performed, each corresponding to a rotation around one of the axes. To represent rotation, we use an x-y-z (or pitch-roll-yaw) *convention*:

$$R^{3D} = R^{3D}(\gamma_x, \gamma_y, \gamma_z) = R_x(\gamma_x)\, R_y(\gamma_y)\, R_z(\gamma_z), \tag{B.1.2}$$

where γ_x, γ_y, γ_z are the three rotation angles, and R_x, R_y and R_z are the corresponding rotational matrices. There is a difference in matrix definition compared to the x-y-z convention, as angle γ_y has the opposite positive direction[1] (see Figure 2.5), which corresponds to the IEC gantry angle definition [5] used in the XVI. Hence, the sine signs in R_y are switched, and the three rotational matrices are:

$$R_x(\gamma_x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\gamma_x & \sin\gamma_x & 0 \\ 0 & -\sin\gamma_x & \cos\gamma_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{B.1.3}$$

[1] In the x-y-z convention, all rotation angles increase counter-clockwise if viewed from the negative direction of the corresponding axes.

$$R_y(\gamma_y) = \begin{pmatrix} \cos\gamma_y & 0 & \sin\gamma_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\gamma_y & 0 & \cos\gamma_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{B.1.4}$$

$$R_z(\gamma_z) = \begin{pmatrix} \cos\gamma_z & \sin\gamma_z & 0 & 0 \\ -\sin\gamma_z & \cos\gamma_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{B.1.5}$$

and the matrix of translation is defined as:

$$T^{3D} = T^{3D}(t) = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{B.1.6}$$

B.2 2D Rigid Geometric Transformation in $\bar{\mathrm{U}}\bar{\mathrm{V}}$ and UV Systems

$\bar{\mathrm{U}}\bar{\mathrm{V}}$ and UV are 2D *left-handed* coordinate systems associated with the flat-panel detector $\bar{\mathrm{H}}$ and isocentric plane H, respectively (see Figure 2.5). They are used for the x-ray projections and fluoroscopic images, where the positions of the grid points are either expressed in pixels (integers, $\bar{\mathrm{U}}\bar{\mathrm{V}}$) or millimeters (UV). An application of the 2D rigid transformation $\mathcal{T}^{2D} : \mathbb{R}^3 \times \mathbb{R}^2 \to \mathbb{R}^2$ to a point $p \in \mathbb{R}^2$ results in:

$$q = \mathcal{T}^{2D}(b, p), \text{ or } \tilde{q} = R^{2D}T^{2D}\tilde{p}, \tag{B.2.1}$$

where $b = (t_u, t_v, \gamma)^T$ is a vector of transformation parameters (two shifts and one rotation angle), and 3×3 matrices of 2D rotation and translation R^{2D} and T^{2D} are defined as:

$$R^{2D} = R^{2D}(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{B.2.2}$$

and

$$T^{2D} = T^{2D}(t) = \begin{pmatrix} 1 & 0 & t_u \\ 0 & 1 & t_v \\ 0 & 0 & 1 \end{pmatrix}. \tag{B.2.3}$$

B.3 Inverse Transformations

The inverse of the translation transformation is defined as:

$$(T(t))^{-1} = T(-t), \tag{B.3.1}$$

and the inverse of the rotation is its transpose:

$$R^{-1} = R^T. \tag{B.3.2}$$

Note that if the matrix R is formed as a multiplication of several rotational matrices, its transpose is obtained as:

$$R^T = (R_1 \dots R_k)^T = R_k^T \dots R_1^T. \tag{B.3.3}$$

In general, an inverse of a rigid geometric transformation represented as a multiplication of elementary rotations and translations A_i is the multiplication of the inverses of elementary transformations in the reverse order, i.e.,

$$(A_1 \dots A_k)^{-1} = A_k^{-1} \dots A_1^{-1}. \tag{B.3.4}$$

B.4 Transformation from $\bar{U}\bar{V}$ to UV

To estimate the motion of the fiducial markers from the fluoroscopic images, grid points of which are positioned in the $\bar{U}\bar{V}$ coordinate system with units in pixels, it is necessary to transform the coordinates to the UV system in order to express the motion in millimeters to make it meaningful in terms of the XYZ coordinate system used for patient positioning.

Due to a significant weight of the x-ray tube and detector and their distance from the isocentre, the system is subject to a *flex* during gantry rotation [30, 92]. It means that the point located at the isocentre is not necessarily imaged in the middle of the flat-panel detector[2] but may be displaced by a vector $(u_{\text{off}}, v_{\text{off}})^T$ (see Figure B.1), measured in millimeters. In addition, the detector can deliberately be shifted to control the size of a reconstructed CBCT volume. Therefore, the system needs to be calibrated. The details of the calibration process are explained by Jaffray *et al.* [92].

[2]The centre of our 1024×1024 pixels detector is a middle point between the four central pixels.

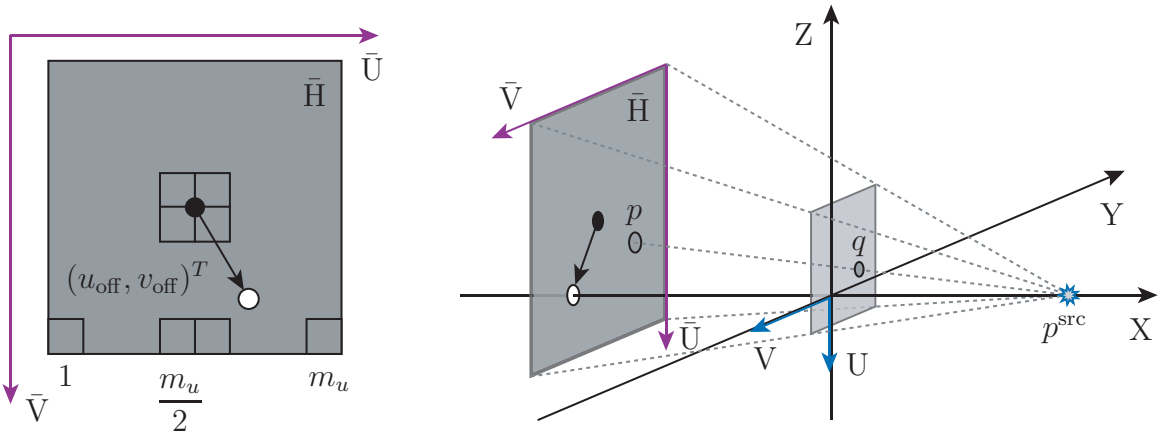


Figure B.1: Piercing point (*white*) is a point of projection of the isocentre on the flat-panel detector plane $\bar{\text{H}}$ with p^{src} centre of projection. The displacement vector $(u_{\text{off}}, v_{\text{off}})^T$ measured in millimeters between the centre of the flat-panel detector (*black point*) and the piercing point change while the LINAC rotates and can be known from calibration.

In short, a ball bearing is placed at the isocentre as represented by the in-room laser system. The bearing centroid, which is called a *piercing point*, is computed from every CBCT projection image, acquired as the gantry performs its rotation. The parameters $(u^j_{\text{off}}, v^j_{\text{off}})^T$ and their corresponding gantry angles ϕ_j are stored to a file called a *calibration map*. The motion of the system is characterized by *hysteresis*, therefore, it matters, from which angle to start, and in what direction, clockwise or counterclockwise, to go. In clinical practice, several calibration maps are created for every LINAC, each for a different imaging protocol. Flex corrections on the order of 2 mm were reported for a similar system [92], with a slightly larger source to detector distance $\chi_{\text{SDD}} = 1,550$ mm versus our setup of 1,536 mm.

Later, before the CBCT image reconstruction can be performed for a patient, all 2D CBCT projection images are corrected by using the flex parameters computed by interpolation from the corresponding calibration map.

However, with the clinical setup of our imaging system, the flex parameters were not recorded for the fluoroscopic images. In fact, there is a possibility that the fluoroscopic sequences need a separate calibration, as the gantry is not moving, and thus the shift associated with the flex may be different comparing to the CBCT projections acquired at the same angle. To overcome the situation where patient

fluoroscopic sequences do not have corresponding calibration maps, we estimate the flex parameters from the parameters of the CBCT projections[3] acquired during the same treatment session as the fluoroscopic sequence (daily setup verification CBCT). Although it may not be extremely accurate, it still provides a reasonable approximation, which is then further improved by the 2D-3D image registration between the CBCT and the fluoroscopic image. The registration also accounts for the organ motion between the CBCT and fluoroscopic image acquisitions.

The flat-panel detector of 1024×1024 points was used to acquire images for this thesis. However, often the images can be downsampled to save space by selecting an appropriate downsampling factor in the XVI software. Often, the CBCT projections are acquired on a 512×512 grid. Let $\tilde{m} = (\tilde{m}_u, \tilde{m}_v)^T$ be the physical size of the detector grid, $m = (m_u, m_v)^T$ be the size of the image grid, and $\tilde{\delta} = (\tilde{\delta}_u, \tilde{\delta}_v)^T$ be a size of the pixel at the detector, in mm. Then, the size of the image pixel can be computed as:

$$\delta_{\text{px}} = \tilde{\delta}_u \frac{\tilde{m}_u}{m_u}. \tag{B.4.1}$$

In our case, the grids are regular and square. However, if the grid is irregular and/or rectangular with $m_u \neq m_v$, other arrangements must be made.

The pixels are numbered starting from 1 for the convenience of implementation, as this is the convention for the Matlab arrays[4]. The transformation from $\bar{\text{U}}\bar{\text{V}}$ to UV involves rescaling and shift of the original pixel grid, and the measurement point $q \in \mathbb{R}^2$ in UV shown in Figure B.1 (*right*) can be computed from the image point $p \in \mathbb{R}^2$ in $\bar{\text{U}}\bar{\text{V}}$ coordinate system as follows:

$$q = \left((p - \frac{m}{2} - 0.5)\, \delta_{\text{px}} - (u_{\text{off}}, v_{\text{off}})^T \right) \frac{\chi_{\text{SAD}}}{\chi_{\text{SDD}}}, \tag{B.4.2}$$

where χ_{SAD} the distance between the x-ray source and the isocentric plane (source to axis distance), and χ_{SDD} is the distance between the x-ray source and the detector plane.

The transformation of the fluoroscopic image grids from the $\bar{\text{U}}\bar{\text{V}}$ to the UV coordinate system is performed in the beginning, before any further processing takes

[3] While the possibility to estimate the flex parameters from the CBCT calibration map is incorporated into Gryphon, we did not have calibration maps for all data sets. Therefore, in order to standardize data processing in this thesis, we always estimate them from the CBCT projections.

[4] The modification of the numbering to start from 0, which may be more convenient in other implementation systems, such as C/C++/OpenCV is straightforward.

Table B.1: Geometric setup parameters.

Parameter	Notation	Value
Source to axis distance	χ_{SAD}	1000 mm
Source to detector distance	χ_{SDD}	1536 mm
Detector element size	$\tilde{\delta}_u$, $\tilde{\delta}_v$	0.4×0.4 mm^2
Detector grid size	$\tilde{m}_u$, $\tilde{m}_v$	1024×1024 px^2
Pixel size at the detector plane ($\bar{\text{U}}\bar{\text{V}}$)	δ_{px}	
1024×1024 projections		0.4×0.4 mm^2
512×512 projections		0.8×0.8 mm^2
Pixel size at the isocentric plane (UV)		
1024×1024 projections		$\sim 0.2604 \times 0.2604$ mm^2
512×512 projections		$\sim 0.5208 \times 0.5208$ mm^2

place, i.e., all models and algorithms presented in this thesis assume that the positions of the grid points are in the UV coordinates, expressed in millimeters. The parameters of our geometric setup are summarized in Table B.1.

B.5 Perspective Transformation from XYZ to UV

In order to compute a selective DRR that is used as the 2D template image in the process of 2D-3D image registration between a 3D CBCT and a fluoroscopic image, a geometric relationship between the XYZ and UV coordinate systems has to be established. Figure 2.5 shows a point $p^0 \in \mathbb{R}^3$. It has to be projected onto the plane H (with p^{src} being the centre of projection) to obtain $p^{\text{iso}} \in \mathbb{R}^3$. Then, the coordinates of p^{iso} have to be expressed in terms of the UV coordinate system as $q \in \mathbb{R}^2$.

Figure 2.5 shows the system in the position in which the kV x-ray source p^{src} is located on the positive part of the X axis, and the isocentric and flat-panel detector planes, H and $\bar{\text{H}}$, respectively, are parallel to the YZ plane, which we call the initial position. However, our 2D fluoroscopic image sequences can be acquired with a different configuration of the imaging system for which p^{src} and $\bar{\text{H}}$ are rotated around the Y axis onto a known angle ϕ. Expressing the point p^0 in homogeneous coordinates is followed by the rotation to the initial position:

$$\tilde{p}^* = R_y(-\phi)\,\tilde{p}^0. \tag{B.5.1}$$

The distance from the isocentre to the x-ray source $\chi_{\text{SAD}} = \|p^{\text{src}}\|$ is known from the technical setup (see Table B.1). The projected point is then computed as:

$$\tilde{p}^{\text{iso}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\frac{1}{\chi_{\text{SAD}}} & 0 & 0 & 1 \end{pmatrix} \tilde{p}^{*}. \tag{B.5.2}$$

Note that after the perspective transformation (B.5.2) the last homogeneous coordinate of $\tilde{p}^{\text{iso}}$ is likely to be different from 1. It is not equal to 0 either, as the measured object is assumed to be situated between the x-ray source and the detector, i.e., the X coordinates of the points that model the object are less than χ_{SAD}. The homogeneous coordinates are discarded by diving $\tilde{p}^{\text{iso}}$ by its last coordinate to obtain p^{iso}. Finally, $p^{\text{iso}} \in \mathbb{R}^3$ is expressed in terms of the UV coordinate system as point $q \in R^2$:

$$q = \begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} p^{\text{iso}} = \begin{pmatrix} -p_z^{\text{iso}} \\ -p_y^{\text{iso}} \end{pmatrix}. \tag{B.5.3}$$

For convenience of referencing and formulations, the operations performed in Equations (B.5.1) to (B.5.3) and accompanying explanations are combined into one operator $\mathcal{P} : \mathbb{R}^3 \to \mathbb{R}^2$:

$$q = \mathcal{P}(p). \tag{B.5.4}$$

Sometimes, it is necessary to estimate the size of a projection of a 3D object, such as the marker or the voxel, to the isocentric plane H. It is assumed that the object is located in the XYZ system close to some point $p^0 \in \mathbb{R}^3$. Let us denote the size of the object by δ^{3D}, and the size of its measurement in the 2D projection image by δ^{2D}. Then, δ^{2D} can be computed as:

$$\delta^{\text{2D}} = \kappa\, \delta^{\text{3D}}, \tag{B.5.5}$$

where

$$\kappa = \frac{\chi_{\text{SAD}}}{\chi_{\text{SAD}} - \tilde{p}_x^{*}}, \tag{B.5.6}$$

and $\tilde{p}^{*}$ is defined in Equation (B.5.1).

Appendix C

Detailed Formulations

C.1 Filter Based on Karhunen-Loève Transform

One of the methods considered in this thesis for temporal filtering (Section 4.1.1) is based on *Karhunen-Loève transform* (KLT) [51, 227]. The detailed description of the KLT can be found in [51] but we summarize it here for the convenience of the reader. For a sequence of discrete images, $\mathtt{S}_0, \ldots, \mathtt{S}_{m-1}$, a matrix A is constructed so that row i corresponds to $\mathrm{vec}(\mathtt{S}_i)^T$, where $\mathrm{vec}(\cdot)$ is the column stacking operator, i.e., $\mathrm{vec}(\mathtt{S}_i)$ is created from image $\mathtt{S}_i$ by stacking its columns one by one to form a long vector. Let n denote the length of $\mathrm{vec}(\mathtt{S}_i)^T$. The $m \times m$ temporal covariance matrix computed as AA^T/n has m eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m$ and a corresponding eigenvector matrix E with eigenvectors as rows, which is called the KLT matrix. The matrix $B = EA$ contains *eigenimages* represented as rows and sorted in the descending order in accordance with the eigenvalues. The filtered image matrix $\bar{A}$ is then reconstructed from the first $k \leq m$ eigenimages as $\bar{A} = E_k^T B_k$, where E_k and B_k are the matrices composed of the first k rows of E and B, respectively. Finally, the filtered images are produced by reshaping the rows of $\bar{A}$ into the 2D arrays $(\bar{\mathtt{S}}_i)$, $i = 0, \ldots, m-1$. Naturally, when $k = m$, the original images are reconstructed. A discussion on the choice of k is provided in Section 4.1.3. From an implementation perspective, the KLT method conceptually differs from the recursive and FIR filtering by its ability to produce several filtered images simultaneously. Hence, in addition to a choice of k, another design decision that should be made is whether the KLT should be applied to the whole sequence or whether, and how, the latter should be separated into intersecting or non-intersecting subsequences, and how the results

should be combined.

C.2 Finite Impulse Response Filters

One of the approaches for the temporal filtering of the fluoroscopic sequences is the use of finite impulse response (FIR) filters, discussed in Section 4.1.1. Let $N = \ell - 1$, where ℓ is a filter kernel size, as defined in Section 4.1.1. The CNR results presented in Section 4.1.3 were produced by the application of the following filters to the "MM 23" data set:

- **Dolph-Chebyshev** window is defined in terms of samples of its Fourier transform [80, 1]:

$$\tilde{W}(k) = (-1)^k \frac{\cos\left(N \cos^{-1}\left(\beta \cos\left(\frac{\pi k}{N}\right)\right)\right)}{\cosh\left(N \cosh^{-1}(\beta)\right)}, \tag{C.2.1}$$

  where $0 \leq |k| \leq N-1$, $\beta = \cosh\left(\frac{1}{\ell} \cosh^{-1}(10^{\alpha})\right)$. To obtain the corresponding window time samples $\tilde{w}_n$, a discrete Fourier transform is performed on the samples $\tilde{W}(k)$ and then scaled to unity peak amplitude. Parameter α is the logarithm of the ratio of main-lobe level to side-lobe level.
- **Flat top** is a summation of cosines [1]:

$$\tilde{w}_n = \sum_{k=0}^{4} (-1)^k a_k \cos\left(2k \frac{\pi n}{N}\right), \tag{C.2.2}$$

  where $0 \leq n \leq N$, and the coefficient values are $a_0 = 0.21557895$, $a_1 = 0.41663158$, $a_2 = 0.277263158$, $a_3 = 0.083578947$, $a_4 = 0.006947368$.
- **Gaussian** filter coefficients are computed as [80, 1]:

$$\tilde{w}_n = \exp\left(-\frac{1}{2}\left(\frac{n - N/2}{\sigma N/2}\right)^2\right). \tag{C.2.3}$$

  Parameter $0 < \sigma \leq 0.5$ is the Gaussian's standard deviation.
- **Hamming** filter [80, 32, 1]:

$$\tilde{w}_n = 0.54 + 0.46 \cos\left(\frac{2\pi(n - N/2)}{N}\right). \tag{C.2.4}$$

- **Mean**, or rectangular, filter [80, 32, 1]:

$$\tilde{w}_n = 1. \tag{C.2.5}$$

- **Triangular** filter for ℓ odd [1]:

$$\tilde{w}_n = \begin{cases} \dfrac{2(n+1)}{\ell+1}, & 0 \le n \le \dfrac{\ell-1}{2}, \\ \dfrac{2(\ell-n)}{\ell+1}, & \dfrac{\ell+1}{2} \le n \le \ell-1, \end{cases} \tag{C.2.6}$$

and for ℓ even:

$$\tilde{w}_n = \begin{cases} \dfrac{2n+1}{\ell}, & 0 \le n \le \dfrac{\ell}{2}-1, \\ \dfrac{2(\ell-n)-1}{\ell}, & \dfrac{\ell}{2} \le n \le \ell-1. \end{cases} \tag{C.2.7}$$

- **Tukey** window is a constructed window that can be described as a cosine lobe of width $(\alpha/2)N$ convolved with a rectangle window of width $(1-\alpha/2)N$ [80, 1]:

$$\tilde{w}_n = \begin{cases} \beta(n), & 0 \le n \le (1-\alpha)\,\dfrac{N}{2}, \\ 1, & (1-\alpha)\,\dfrac{N}{2} \le n \le (1+\alpha)\,\dfrac{N}{2}, \\ \beta(n), & (1+\alpha)\,\dfrac{N}{2} \le n \le N, \end{cases} \tag{C.2.8}$$

where

$$\beta(n) = \frac{1}{2}\left(1 + \cos\left(\pi\,\frac{n-(1+\alpha)N/2}{(1-\alpha)N}\right)\right) \tag{C.2.9}$$

and $0 \le \alpha \le 1$ influences the widths of the window's central rectangular and side cosine parts.

To find $\tilde{w}_n$, we use standard Matlab implementations[1] `chebwin`, `flattopwin`, `gausswin`, `hamming`, `rectwin`, `triang`, `tukeywin`. After obtaining coefficients $\tilde{w}_n$, they are normalized:

$$w_n = \frac{\tilde{w}_n}{\sum_{n=0}^{\ell-1} \tilde{w}_n}. \tag{C.2.10}$$

To provide a visual example, the normalized window functions of all FIR filters described above for $\ell = 201$ are shown in Figure C.1.

[1]The implementation of all FIR filters presented in this section is from the Signal Processing Toolbox of Matlab R2009b (The Mathworks, Inc., Natick, MA) including default values for the filter parameters if any.

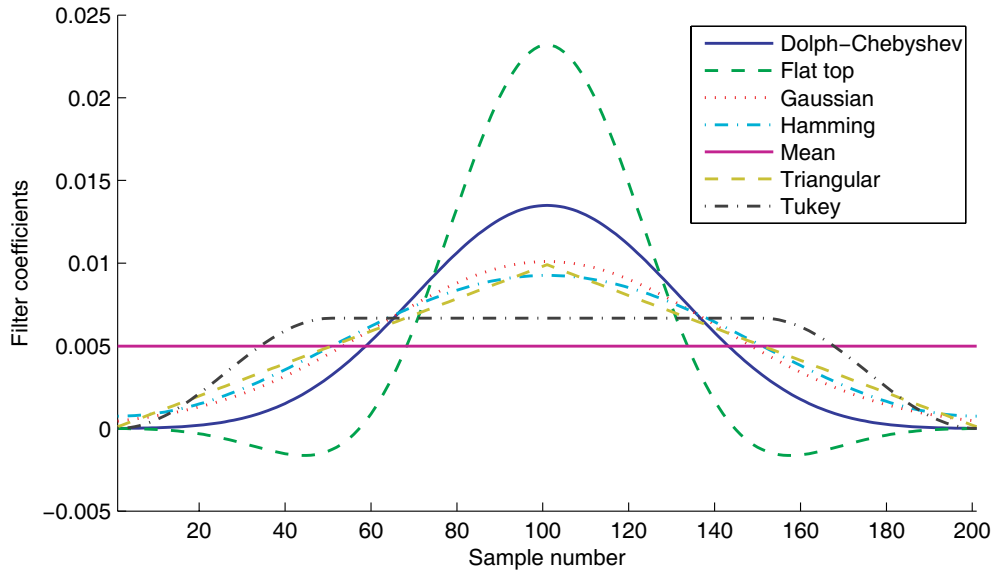


Figure C.1: Normalized window functions for Dolph-Chebyshev, flat top, Gaussian, Hamming, mean, triangular, and Tukey filters for $\ell = 201$.

C.3 Raytracing for Marker Models

Cylindrical Marker Models

The following discussion provides formulations for a marker model, $\mathrm{M}^{\mathrm{mod}}$, used for marker model selective DRR computation described in Section 5.1.1. The extension to multiple marker models is straightforward. Let the points $c_{3\mathrm{D}}$, e^1, and e^2 be the centre and endpoints of $\mathrm{M}^{\mathrm{mod}}$, respectively, found as explained in Section 3.4. Before the 2D template image can be generated, the cylinder's transformation has to be found, i.e., the position and orientation of $\mathrm{M}^{\mathrm{mod}}$ has to be expressed as the 3D rigid transformation that, when applied to a cylindrical model centered at the XYZ origin whose axis of symmetry is aligned with the Z axis, gives $\mathrm{M}^{\mathrm{mod}}$. This transformation can be computed from the points $c_{3\mathrm{D}}$, e^1, and e^2 as explained below.

In the following, we operate on 3D geometrical objects, however, all points are expressed in homogeneous coordinates (explained in Appendix B), i.e., the points are in $\mathbb{R}^4$ and the transformations are the matrices in $\mathbb{R}^{4\times4}$. The cylindrical surface of

radius r aligned with the Z axis is defined as a set of points s, such that:

$$s^T Q s = 0, \tag{C.3.1}$$

where Q is given as:

$$Q = \begin{pmatrix} \frac{1}{r^2} & 0 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{C.3.2}$$

In order to bring the cylindrical surface to a given position and orientation (to describe the positions and orientations of the markers), the matrix Q in Equation (C.3.1) has to be substituted by:

$$A^{\text{cyl}} = F^T Q F, \tag{C.3.3}$$

where F is the 3D rigid geometric transformation computed as follows. The centre $c_{3\text{D}}$ and the endpoints e^1 and e^2 are known from the computation of the 3D marker model discussed in Section 3.4. Then, F has to move the point $c_{3\text{D}}$ to the origin of the XYZ coordinate system, and points e^1 and e^2 to $(0, 0, h/2)^T$ and $(0, 0, -h/2)^T$, where h is the length of the fiducial marker. Such positioning corresponds to a cylinder solid whose axis is aligned with the Z axis, and whose central point coincides with the origin of the XYZ system.

The transformation F consists of a translation and rotations. Since a cylinder is symmetric around its axis, only two rotation angles are required. Thus, we define the transformation matrix as:

$$F = R_x(-\gamma_x)\, R_z(-\gamma_z)\, T^{3\text{D}}(-c_{3\text{D}}). \tag{C.3.4}$$

Appendix B.1 provides formulations for the rotational matrices R_x, R_z and translation $T^{3\text{D}}$. It does not matter which two angles out of γ_x, γ_y and γ_z are selected, however, we prefer to work with γ_x and γ_z as their positive directions in the IEC gantry angle definition [5] used in this thesis coincide with the widely accepted x-y-z convention (more details can be found in Appendix B.1).

Figure C.2 provides an illustration for computing the transformations for the cylindrical marker model. Let $p^{\text{sh}} = T^{3\text{D}}(-c_{3\text{D}})\, e^1$. If $p^{\text{sh}}_x = p^{\text{sh}}_y = 0$, then the marker model is parallel to the Z axis, and $F = T^{3\text{D}}(-c_{3\text{D}})$. If not, we first rotate p^{sh} around the Z axis until $p^{\text{rot}} = R_z(-\gamma_z)\, p^{\text{sh}} \in$ YZ. The angle γ_z is the angle between the

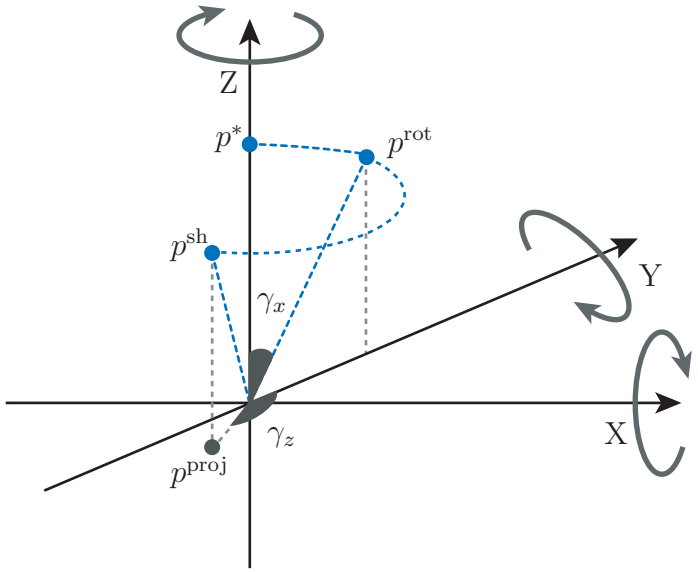


Figure C.2: Transformations for the cylindrical marker model.

projection p^{proj} of the point p^{sh} onto the XY plane and the positive direction of the Y axis, $p^{\text{proj}} = (p_x^{\text{sh}}, p_y^{\text{sh}}, 0, 1)^T$. Then,

$$\sin\gamma_z = \frac{p_x^{\text{sh}}}{\sqrt{(p_x^{\text{sh}})^2 + (p_y^{\text{sh}})^2}}, \quad \cos\gamma_z = \frac{p_y^{\text{sh}}}{\sqrt{(p_x^{\text{sh}})^2 + (p_y^{\text{sh}})^2}}. \tag{C.3.5}$$

Next, we rotate p^{rot} around X axis until $p^* = R_x(-\gamma_x)\, p^{\text{rot}} \in \text{Z}$, where

$$\sin\gamma_x = \frac{p_y^{\text{rot}}}{h/2}, \quad \cos\gamma_x = \frac{p_z^{\text{rot}}}{h/2}. \tag{C.3.6}$$

After the sines and cosines of γ_x and γ_z are found, the matrices R_x and R_z, and then F and A^{cyl} can be computed according to Equations (B.1.3), (B.1.5), (C.3.4), and (C.3.3), respectively. The method of finding A^{cyl} is summarized in Algorithm D.11.

Finally, we define the cylindrical marker model as a set of points s such that:

$$\text{M}^{\text{mod}} = \begin{cases} s^T A^{\text{cyl}}\, s \le 0, \\ (s - e^1)^T (e^1 - c_{\text{3D}}) \le 0, \\ (s - e^2)^T (e^2 - c_{\text{3D}}) \le 0. \end{cases} \tag{C.3.7}$$

The first inequality in (C.3.7) is an infinite cylinder, the second and third are the 3D half-spaces created by the parallel planes perpendicular to the cylinder's axis and going through the endpoints e^1 and e^2, respectively (see Figure C.3). Both half-spaces include the point c_{3D}.

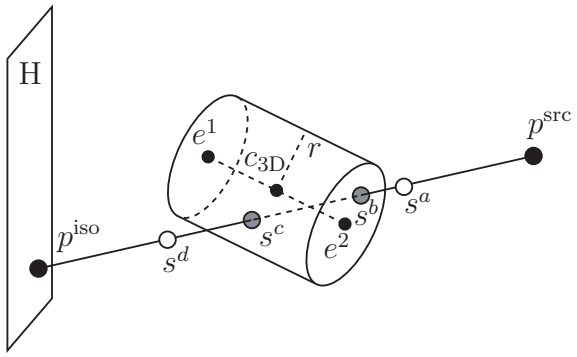


Figure C.3: Raytracing for the cylindrical marker model.

As explained in Section 5.1.1 and shown in Figure C.3, MM DRR generation requires computing the distance that the ray, that goes from the x-ray source, p^{src}, to the point on the detector, p^{iso}, travels inside the marker model M^{mod}. This can be computed by finding the distance between the points of intersection between the ray and M^{mod} by analytically solving the system of equations describing the marker model (C.3.7) and the ray. In general, if the ray does not belong to any of the planes or the cylinder surface, four points of intersection are possible: two with the infinite cylinder surface and one with each of the planes. For the example in Figure C.3, the four points are:

- s^a: with a continuation of the cylinder;
- s^b: with the right plane;
- s^c: with the cylinder;
- s^d: with a continuation of the left plane.

Only two of these points, namely s^b and s^c, are the actual points of intersection of the ray with the marker model. Therefore, for each found point of intersection, it is necessary to verify whether the given point belongs to the marker model described by (C.3.7). In fact, the substitution of each intersection point in place of s in (C.3.7) should turn the left-hand-sides of one or more inequalities into a value that is very close to 0 and the rest to the values less than 0. For example, the point s^b will turn the left-hand-sides of the first and second inequalities into < 0 and the third into ≈ 0. The substitution of the point s^c will result in ≈ 0 for the first inequality and < 0 for the second and third. To have a unified testing procedure for all the cases,

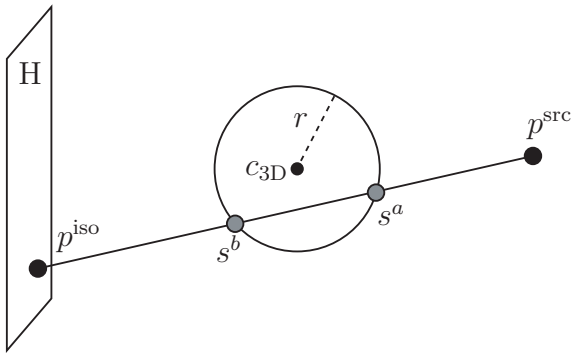


Figure C.4: Raytracing for the spherical marker model.

and to allow for some computational inaccuracy, we require all left-hand-sides of the inequalities in (C.3.7) to be smaller or equal to 10^{-8} instead of 0. This number was found empirically.

For efficiency, the computation described above is only performed for the points on the plane H (centres of the pixels of the 2D template image) that have the potential to contain the "projection" of the 3D marker. The implementation details are provided in Appendix D.3, and summarized in Algorithms D.10 and D.11.

Spherical Marker Models

Let the point c_{3D} be the centre of the marker model $\mathrm{M}^{\mathrm{mod}}$, and let r be its radius (see Figure C.4). After switching to homogeneous coordinates, the spherical marker model can be defined as a set of points s such that:

$$\mathrm{M}^{\mathrm{mod}} = \{s \mid s^T A^{\mathrm{sph}} s \leq 0\}, \tag{C.3.8}$$

where

$$A^{\mathrm{sph}} = F^T Q F. \tag{C.3.9}$$

The matrix Q describes a spherical shape of radius r:

$$Q = \begin{pmatrix} \frac{1}{r^2} & 0 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{C.3.10}$$

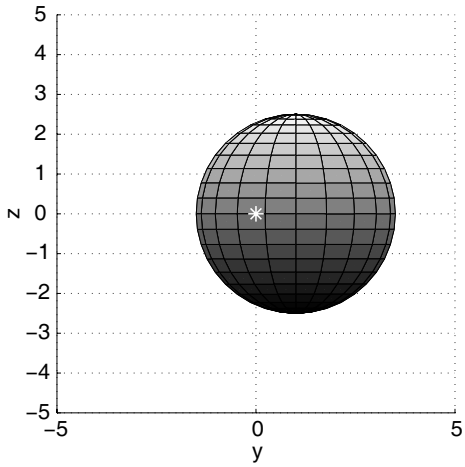


Figure C.5: Spherical model.

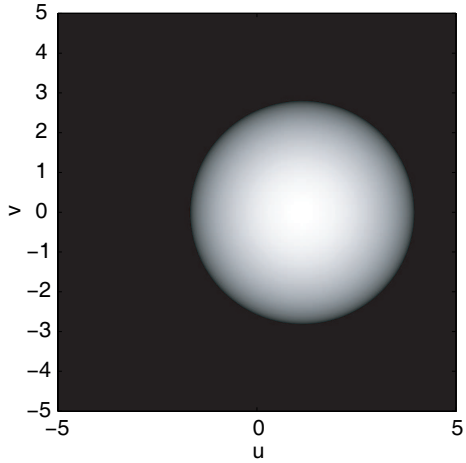


Figure C.6: Template image.

and the matrix F represents a rigid geometric transformation that moves the point c_{3D} to the origin of the XYZ coordinate system, i.e., the translation:

$$F = T^{3D}(-c_{3D}), \tag{C.3.11}$$

where T^{3D} is defined in Appendix B.1.

Similarly to the cylindrical model, the intensity at a point p^{iso} on the isocentric plane H is computed as the distance between the points of intersection between the marker model (C.3.8) and the ray from p^{src} to p^{iso}, namely the points s^a and s^b shown in Figure C.4. This computation is only performed for the rays that have a potential to intersect the marker model.

The implementation details are provided in Appendix D.3, and summarized in Algorithm D.12. If required, the amplification of the marker model size can be performed preceding image generation analogically to the method described for the cylindrical marker models in Section 5.1.1.

Figure C.5 demonstrates a spherical marker model of radius $r = 2.5$, located at $c_{3D} = (2, 1, 0)^T$, and Figure C.6 its integral projection on the H plane. Centre of projection in Figure C.5 is marked by a *star*. It coincides with the origin of the XYZ coordinate system.

C.4 Distance Measures

This sections provides a review of commonly used distance measure functions used in image registration. To remind, the distance measure $\mathcal{D}$ is a function that quantifies the difference between two images by a real number. Let $\mathtt{A}, \mathtt{B} : \Gamma^2 \cap \Omega_{\text{ROI}} \to \mathbb{R}$ be the two discrete images to register, where $\Gamma^2 \subset \mathbb{R}^2$ is the pixel grid, and Ω_{ROI} is a registration ROI.

Intensity difference

One of the measures often used in single-modality registration is the *sum of squared intensity differences (SSD)*:

$$\mathcal{D}_{\text{SSD}}(\mathtt{A}, \mathtt{B}) = \sum_{i=1}^{N} \left(\mathtt{A}[\check{p}^i] - \mathtt{B}[\check{p}^i]\right)^2, \tag{C.4.1}$$

where N is a number of grid points $\check{p}^i$ in Ω_{ROI}. Examples of the SSD application include 2D-3D registration between digital subtraction angiographic (DSA) and magnetic resonance angiographic (MRA) images [41], the preliminary step of 2D-3D registration of the x-ray and CT images in cranial radiosurgery [64], and serial MR registration [76].

The SSD is very sensitive to a small number of pixels with significantly different intensities, e.g., in the case when a contrast material is injected to the patient between the acquisition of the images $\mathtt{A}$ and $\mathtt{B}$ [83]. This can be partially reduced by using the *sum of absolute differences (SAD)* instead of the SSD:

$$\mathcal{D}_{\text{SAD}}(\mathtt{A}, \mathtt{B}) = \sum_{i=1}^{N} \left|\mathtt{A}[\check{p}^i] - \mathtt{B}[\check{p}^i]\right|. \tag{C.4.2}$$

The SAD was reported to be used for 3D registration of SPECT and PET images [52], and 3D registration of PET images [86].

In the conventional definition, both $\mathcal{D}_{\text{SSD}}$ and $\mathcal{D}_{\text{SAD}}$ are divided by N to account for possible change in the overlap between the images in the process of registration [83]. Since our registration problems are solved over a fixed ROI, and the division by the number of pixels may result in very small distance function values, which may complicate optimization (see Section 5.2), we drop the scaling.

Correlation

The SSD and SAD methods deal directly with intensities and hence they need a direct correspondence of intensities in $\mathtt{A}$ and $\mathtt{B}$ for meaningful registration. Otherwise, the image intensities have to be modified before the registration. In contrast, the *normalized correlation coefficient (NCC)* only assumes a linear relationship between the intensity values in the images, and thus may be more suitable for a wider range of applications:

$$\mathcal{D}_{\mathrm{NNCC}}(\mathtt{A},\mathtt{B}) = -\frac{\sum_{i=1}^{N}\left(\mathtt{A}[\check{p}^i]-\mu_{\mathtt{A}}\right)\left(\mathtt{B}[\check{p}^i]-\mu_{\mathtt{B}}\right)}{\sqrt{\sum_{i=1}^{N}\left(\mathtt{A}[\check{p}^i]-\mu_{\mathtt{A}}\right)^2\sum_{i=1}^{N}\left(\mathtt{B}[\check{p}^i]-\mu_{\mathtt{B}}\right)^2}}, \tag{C.4.3}$$

where $\mu_{\mathtt{A}}$ and $\mu_{\mathtt{B}}$ are the mean intensity values of $\mathtt{A}$ and $\mathtt{B}$, respectively. The formulation (C.4.3) is the negation of the conventional NCC definition given in, e.g., Hill *et al.* [83], as our image registration task (5.0.1) is a minimization problem, and in the literature it is customary to formulate it as the maximization of the NCC. The NCC was successfully used in registration of CBCT to fluoroscopic images based on the measurements of the long coil markers for prostate motion evaluation [7], estimation of 3D orientation of the CT volume of a vertebra from its 2D fluoroscopic images [26], patient C-arm pose estimation by registration of 2D C-arm images to 3D ultrasound [54], and detection of the markers in 2D kV or MV images for tracking applications by maximizing the correlation with the template image [43, 59, 169, 174].

Information Theoretic Techniques

The distance measures derived from the concepts of information theory rely on the idea that the amount of information in the combined image should reduce when the images are aligned [150, 83, 211]. In brief, the measure of information is the Shannon-Wiener entropy that can be defined as $\mathrm{H} = -\sum_{i=1}^{N} h_i \log h_i$, where H is the average information supplied by N symbols of the discrete finite alphabet whose probabilities of occurrence are $h_1, \ldots, h_N$ [83]. Entropy H is maximized when all symbols have equal probability of occuring, i.e., $h_i = 1/N$, and minimized (equal to 0) if the probability of one symbol occurring is 1, and all the others is 0. In the application to image

registration, the lower the entropy of the combined image, the better is the alignment between the images.

Some of the popular distance measures based on the entropy are explained below. The marginal entropy of the image $\mathtt{A}$ is defined as [83]:

$$\mathrm{H}(\mathtt{A}) = -\sum_{a} h_{\mathtt{A}}(a) \log h_{\mathtt{A}}(a), \tag{C.4.4}$$

where $h_{\mathtt{A}}(a)$ is the probability that the points (pixels) in image $\mathtt{A}$ have the intensity value a. The value $\mathrm{H}(\mathtt{A})$ can be calculated from the (normalized version of the) image intensity histogram, in which the probabilities $h_{\mathtt{A}}(a_1), \ldots, h_{\mathtt{A}}(a_k)$ are the histogram entries. The value $\mathrm{H}(\mathtt{B})$ is computed by analogy. The joint entropy is defined as [83]:

$$\mathrm{H}(\mathtt{A}, \mathtt{B}) = -\sum_{a} \sum_{b} h_{\mathtt{A},\mathtt{B}}(a, b) \log h_{\mathtt{A},\mathtt{B}}(a, b), \tag{C.4.5}$$

where $h_{\mathtt{A},\mathtt{B}}(a, b)$ correspond to the probability of pairs of image values a and b occurring in the corresponding points of $\mathtt{A}$ and $\mathtt{B}$. Modersitzki [150] provides a detailed discussion on estimating the joint density $h_{\mathtt{A},\mathtt{B}}(a, b)$.

One of the common registration techniques is the maximization of the *mutual information (MI)*, where the best alignment of the images is considered when a low joint entropy $\mathrm{H}(\mathtt{A}, \mathtt{B})$ and high image entropies $\mathrm{H}(\mathtt{A})$ and $\mathrm{H}(\mathtt{B})$ are achieved. The distance measure based on the MI [83] is defined as:

$$\mathcal{D}_{\mathrm{MI}}(\mathtt{A}, \mathtt{B}) = -\left(\mathrm{H}(\mathtt{A}) + \mathrm{H}(\mathtt{B}) - \mathrm{H}(\mathtt{A}, \mathtt{B})\right), \tag{C.4.6}$$

where the negation is due to minimization of the distance measure in (5.0.1). The MI was independently proposed by Collignon *et al.* [46] and Viola [225] for multi-modality medical image registration. It was successfully used in such applications as determination of the pose of an object in 3D from its x-ray projections [13, 225], 2D-2D registration between the x-ray projections and the DRRs generated from the CT data sets [29], 2D-3D registration between x-ray/fluoroscopic and CT images [102, 249], and CT to MRI registration of the brain images [135].

The formulation (C.4.6) has been shown to be sensitive to changes in the amount of overlap between the reference and template images. In particular, changes in overlap of very low-intensity regions, such as air around the patient, can disproportionately contribute to the MI value [211, 70, 83]. Hance, the *normalized MI (NMI)*

was introduced [211]. The distance measure based on NMI is defined as:

$$\mathcal{D}_{\mathrm{NMI}}(\mathtt{A},\mathtt{B}) = -\frac{\mathrm{H}(\mathtt{A})+\mathrm{H}(\mathtt{B})}{\mathrm{H}(\mathtt{A},\mathtt{B})}, \tag{C.4.7}$$

where the negation is due to minimization in (5.0.1). The examples of using NMI include alignment of MR, CT and PET images of the brain [211, 118], and registration of CT DRRs to x-ray or fluoroscopic images of the phantom and chest [42] or the liver [27].

Pattern Intensity

To align the fluoroscopic images obtained during surgical interventions to the pre-treatment CT data set, a new distance measure called *Pattern Intensity (PI)* was proposed by Weese *et al.* [231] and Penney *et al.* [172]. It is based on the idea that if the DRR generated from the object of interest, such as a segmented vertebra, is subtracted from the x-ray image, the vertebra structures will diminish in the difference image when a proper alignment is achieved. The PI considers that a pixel belongs to a structure if it has a significantly different intensity value from the neighbouring pixels. The difference image is defined as:

$$\mathtt{C} = \mathtt{A} - \kappa\mathtt{B}, \tag{C.4.8}$$

where κ is an intensity scaling factor. For every pixel $\check{p}$, its intensity is compared to the intensities of the points $\check{s}^i$, $i = 1, 2, \ldots$, in its neighbourhood, such that the distance between $\check{p}$ and $\check{s}^i$ is less or equal to r. In [231] and [172], $r = 3$ pixels is suggested. The distance measure based on the PI is defined as:

$$\mathcal{D}_{\mathrm{PI}}(\mathtt{A},\mathtt{B}) = -\sum_{\check{p}} \sum_{\check{s}^i}^{\|\check{p}-\check{s}^i\|\leq r} \frac{\alpha^2}{\alpha^2 + \left(\mathtt{C}[\check{p}] - \mathtt{C}[\check{s}^i]\right)^2}, \tag{C.4.9}$$

where α is the sensitivity of the PI to image structures, which should be selected so that it is larger than the standard deviation of the noise in the x-ray/fluoroscopic images but smaller than the contrast of features of interest. As such, it depends on the intensity range of the image. For efficient computation and meaningful registration, it is suggested that the DRR is only generated from the segmented points, and the PI is only computed for non-zero DRR points and their respective neighbourhoods [231].

The PI was used for registration of fluoroscopic and CT spinal images [231, 172], and 2D-3D registration of CT and simulated x-ray images of the tibia, pelvis and skull base [29]. It is claimed that the use of the PI results in stable and accurate registration even in the presence of interventional instruments in one of the images and soft-tissue structures that can reduce the CNR significantly [172].

In the experiment described in Section 5.2, we only compute $\mathcal{D}_{\text{PI}}$ for non-zero DRR points and their neighbours, and use the empirically selected parameters $r = 3$ pixels and $\alpha = 0.1$.

Gradient Difference

The gradient difference method operates on image gradients computed along the coordinate directions. To build insight, let us assume that the images are continuous. In this case, the gradients of the difference image are defined as [172]:

$$\mathtt{H}_j = \frac{\partial \mathtt{A}}{\partial p_j} - \kappa \frac{\partial \mathtt{B}}{\partial p_j}, \tag{C.4.10}$$

where $j = \{u, v\}$, and κ is an intensity scaling factor. In practice, finite differences method can be used to approximate gradients. The distance measure based on the *gradient difference (GD)* [172] is computed as:

$$\mathcal{D}_{\text{GD}}(\mathtt{A}, \mathtt{B}) = -\sum_j \sum_{\check{p}} \frac{\theta_j}{\theta_j + (\mathtt{H}_j[\check{p}])^2}, \tag{C.4.11}$$

where θ_j are the sensitivity constants. In [172], it is suggested that θ_j are computed as the variances of the respective gradient images $\mathtt{H}_j$. Similar to the PI, the GD is reported to produce stable and accurate registration results even in the presence of soft tissues and intervention instruments in one of the images [172, 64].

Appendix D

Implementation Details and Algorithm Listings

D.1 3D Marker Segmentation and Modelling

Algorithm D.1 summarizes methods described in Sections 3.1 and 3.2, i.e., thresholding $\langle\!\langle \mathbf{S}_1 \rangle\!\rangle$, clustering $\langle\!\langle \mathbf{S}_2 \rangle\!\rangle$ and candidate marker set selection $\langle\!\langle \mathbf{S}_3 \rangle\!\rangle$ of the marker segmentation procedure. For better understanding of the approach, we provide details on how the clusters are formed in step $\langle\!\langle \mathbf{S}_2 \rangle\!\rangle$ by the method of hierarchical clustering, though in the Gryphon Matlab function `clusterdata` [1] is used in the following format:

`T = clusterdata(`$(\breve{p}^i)$`, 'cutoff',` η`, 'criterion', 'distance');`

where $\breve{p}^i \in \Theta$ and $(\breve{p}^i)$ is a $|\Theta| \times 3$ array of point coordinates. Resulting vector `T` of size $|\Theta|$ contains cluster numbers for the points.

Next, Algorithm D.2 provides implementation details for the step $\langle\!\langle \mathbf{S}_4 \rangle\!\rangle$, the selection of the marker sets among the candidates based on their proximity to the LINAC's isocentre, point p^{prox}.

Finally, Algorithm D.3 gives details for finding the orientations of the marker models $\langle\!\langle \mathbf{S}_5 \rangle\!\rangle$ by the use of the weighted principal component analysis (PCA). First, we shift and weight the points of the marker sets, and then use the Matlab function `princomp` [1]. Let X be a matrix of size $m \times 3$ composed of the weighted points $w_i p^i$, where m is a number of points in a marker set. Then, the first principal direction, $v^{1,k}$, is computed as:

```
coeff = princomp(X); v1k = coeff(:, 1);
```

Algorithm D.1: Segmentation of the candidate marker sets and estimation of the marker positions in the 3D CBCT image.

Input:
$\Omega_{\text{VOI}} \subset \Gamma^3$ – volume of interest,
$\mathtt{V}_{\text{VOI}}$ – VOI in CBCT image,
θ – segmentation intensity threshold,
η – minimal distance between the marker sets,
ϵ_ℓ and ϵ_{u} – lower and upper marker set size margins,
δ_{m} – marker size,
$[\xi_\ell, \xi_{\text{u}}] = [0.1, 1]$ – intensity rescaling range, Sect. 3.2

s_1 **Thresholding:** $\Theta = \{\check{p}^i \in \Omega_{\text{VOI}} \mid \mathtt{V}_{\text{VOI}}[\check{p}^i] \geq \theta,\ i = 1, \ldots, |\Omega_{\text{VOI}}|\}$

s_2 **Clustering:**
create initial clusters $\hat{\text{M}}_1, \ldots, \hat{\text{M}}_q$, $q \leftarrow |\Theta|$ (one point in each cluster)
compute all $\text{dst}(\hat{\text{M}}_k, \hat{\text{M}}_j)$, $k = 1, \ldots, q-1$, $j = k+1, \ldots, q$
$\upsilon \leftarrow \min\{\text{dst}(\hat{\text{M}}_k, \hat{\text{M}}_j)\}$

while $\upsilon < \eta$ **and** $q > 1$ **do**

select $\hat{\text{M}}_{k_1}$ and $\hat{\text{M}}_{k_2}$ such that $\text{dst}(\hat{\text{M}}_{k_1}, \hat{\text{M}}_{k_2}) = \upsilon$
$\hat{\text{M}}_k \leftarrow \hat{\text{M}}_{k_1} \cup \hat{\text{M}}_{k_2}$
discard $\hat{\text{M}}_{k_1}$, $\hat{\text{M}}_{k_2}$, $\text{dst}(\hat{\text{M}}_{k_1}, \cdot)$, and $\text{dst}(\hat{\text{M}}_{k_2}, \cdot)$
$q \leftarrow q - 1$
compute $\text{dst}(\hat{\text{M}}_k, \hat{\text{M}}_\ell)$, $\ell = 1, \ldots, q$, $\ell \neq k$
find $\upsilon \leftarrow \min\{\text{dst}(\hat{\text{M}}_k, \hat{\text{M}}_\ell)\}$

end

s_3 **Candidate marker set selection:**
for $k \leftarrow 1$ **to** q **do**
if $\text{diam}(\hat{\text{M}}_k) < \delta_{\text{m}} - \epsilon_\ell$ **or** $\text{diam}(\hat{\text{M}}_k) > \delta_{\text{m}} + \epsilon_{\text{u}}$ **then**
discard $\hat{\text{M}}_k$
else
compute weights w_i by rescaling $(\mathtt{V}_{\text{VOI}})_i$ to $[\xi_\ell, \xi_{\text{u}}]$ range
for points $\check{p}^i \in \hat{\text{M}}_k$
compute $c^k_{\text{3D}} \leftarrow \text{mean}(\hat{\text{M}}_k, w)$
end
end

Output:
$\hat{\text{M}}_k$, $k = 1 \ldots, \hat{n} \leq q$ – candidate marker sets in 3D,
c^k_{3D} – candidate marker set positions in the 3D CBCT, centres of $\hat{\text{M}}_k$

Algorithm D.2: Correction for the candidate marker set selection procedure based on the estimated marker positions.

Input:
$\hat{\mathrm{M}}_k$, $k = 1, \ldots, \hat{n}$ – candidate marker sets,
$c_{3\mathrm{D}}^k \in \mathbb{R}^3$ – positions of the candidate marker sets,
$p^{\mathrm{prox}} \in \mathbb{R}^3$ – target proximity point,
n – number of the fiducial markers

$\mathbf{s}_4$ **Correction of the candidate marker set selection:**

if $\hat{n} > n$ **then**
  for $k \leftarrow 1$ **to** $\hat{n}$ **do**
    $d_k = \|c_{3\mathrm{D}}^k - p^{\mathrm{prox}}\|$
  end
  Perform ascending sorting for d: $d_{j_1} \leq \ldots \leq d_{j_{\hat{n}}}$
  for $k \leftarrow 1$ **to** n **do**
    $\mathrm{M}_k \leftarrow \hat{\mathrm{M}}_{j_k}$
  end
else
  for $k \leftarrow 1$ **to** $\hat{n}$ **do**
    $\mathrm{M}_k \leftarrow \hat{\mathrm{M}}_k$
  end
  $n \leftarrow \hat{n}$
end

Output: $\mathrm{M}_1, \ldots, \mathrm{M}_n$ – marker sets in 3D

Algorithm D.3: Estimation of orientations for the cylindrical marker models from the marker sets using a weighted PCA.

Input:
M_k, $k = 1, \ldots, n$ – marker sets,
$c^k_{3\mathrm{D}}$ – centres of the marker sets,
$\mathtt{V}_{\mathrm{VOI}}$ – VOI in CBCT image,
$[\xi_\ell, \xi_\mathrm{u}] = [0.1, 1]$ – intensity rescaling range, Sect. 3.4
h – height of the cylindrical marker

$\mathbf{s}_5$ **for** $k \leftarrow 1$ **to** n **do**

Weights:
compute weights w_i by rescaling $\mathtt{V}_{\mathrm{VOI}}[\check{p}^i]$ to $[\xi_\ell, \xi_\mathrm{u}]$ range
for points $\check{p}^i \in \mathrm{M}_k$

Directions:
$\forall \check{p}^i \in \mathrm{M}_k$: $\check{p}^i \leftarrow \check{p}^i - c^k_{3\mathrm{D}}$
perform PCA on set of weighted points $\{w_i \check{p}^i\}$, find first principal direction $v^{1,k}$

Cylinder endpoints:
$e^{1,k} \leftarrow c^k_{3\mathrm{D}} + v^{1,k} h/2$
$e^{2,k} \leftarrow c^k_{3\mathrm{D}} - v^{1,k} h/2$

end

Output: $e^{1,k}, e^{2,k}, k = 1, \ldots, n$ – endpoints of the cylindrical marker models

D.2 Temporal and Marker Enhancement Filtering

In our implementation, in addition to marker localization and tracking, the same image opening procedure is used for other tasks, e.g., image preview. To allow the preview of the temporally filtered images, for example, in order to visually assess the improvement in quality, temporal filtering is built into the image opening procedure, and is applied to the whole image. Hence, the ROI is applied after temporal filtering but before the marker enhancement.

Algorithm D.4 gives details of implementation of temporal filtering $\langle\!\langle \mathbf{F}_1 \rangle\!\rangle$, described in Section 4.1. It was concluded that the Dolph-Chebyshev is the most suitable among considered existing filters for application to the fluoroscopic sequences that require reduction of the noise, and for which marker motion is expected. A Matlab function `chebwin` [1] is used to compute the filter coefficients:

$w \leftarrow$ `chebwin(`ℓ`);` $w \leftarrow w/$`sum(`$\tilde{w}$`);`

Algorithm D.5 provides the details of the computation of the circular marker and rectangular ROIs, Υ_k and Ω_{ROI}, respectively, that were discussed in Section 4.3 and are applied to the image at step $\langle\!\langle \mathbf{F}_2 \rangle\!\rangle$. The ROIs themselves are computed in the localization (step $\langle\!\langle \mathbf{L}_1 \rangle\!\rangle$) and tracking (step $\langle\!\langle \mathbf{T}_2 \rangle\!\rangle$) procedures. The ROI computation is based on the estimation of the maximal possible 2D displacement, ϱ_k, $k = 1, \ldots, n$, from the predicted marker positions, $\hat{c}^k_{\text{2D}}$, that can either be computed as projections of the 3D positions, $\mathcal{P}(c^k_{\text{3D}})$ in 2D-3D image registration of the localization procedure, or from the preceding images of the sequence in 2D-2D registration of the tracking. The values ϱ_k are computed from the geometric transformation, $\mathcal{T}^{\text{2D}}$ or $\mathcal{T}^{\text{3D}}$, and the assumptions on maximal magnitude of marker motion, which are expressed as the bounds on geometric transformation parameters, $|a_j| \leq a^{\max}_j$. The optimization problem of finding ϱ_k is solved using SQP implemented as the active set method in Matlab function `fmincon` [1]. Should the optimization fail or return an unreasonably small value for ϱ_k caused by sticking in the local optimum, the ϱ_k is assigned the maximal displacement computed from the upper bounds on the shift parameters $t^{\max}_u, t^{\max}_v$ or $t^{\max}_x, t^{\max}_y, t^{\max}_z$ that constitute a part of the $a^{\max}$ vector.

Finally, step $\langle\!\langle \mathbf{F}_3 \rangle\!\rangle$ corresponds to the application of the novel marker enhancement filter (MEF), described in Section 4.4, to the ROI in the temporally filtered

image[1] $\bar{\mathtt{S}}$ in order to produce a MEF-image $\mathtt{R}$, that is further used in registration of either localization or tracking procedure. In $\mathtt{R}$, the markers are amplified while anatomical features, background, and the noise are suppressed. Here, we describe the full MEF used in localization. The reduced variation of the MEF, the magnitude-only filter (MOF) used in tracking, can be straightforwardly deduced by selecting appropriate steps from the MEF as described in Section 4.4.5.

As explained in Section 4.4.1, the MEF starts with construction of a scale-space representation $\mathcal{L}$ of $\bar{\mathcal{S}}$, which involves the convolution with the Gaussian of scale σ (Section 4.2.2). The magnitude-only and magnitude-and-ratio images, $\mathcal{Y}_{\mathrm{m}}$ and $\mathcal{Y}_{\mathrm{mr}}$, respectively, are built using the eigenvalues $\lambda_1(p,\sigma)$ and $\lambda_2(p,\sigma)$ of the 2×2 Hessian matrices $H(p,\sigma)$, where the Hessians are constructed from the second derivatives of $\mathcal{L}$, $\mathcal{L}_{ij}(p,\sigma) = \frac{\partial^2}{\partial p_i \partial p_j}\mathcal{L}(p,\sigma)$, $i,j \in \{u,v\}$:

$$H(p,\sigma) = \begin{bmatrix} \mathcal{L}_{uu}(p,\sigma) & \mathcal{L}_{uv}(p,\sigma) \\ \mathcal{L}_{uv}(p,\sigma) & \mathcal{L}_{vv}(p,\sigma) \end{bmatrix}. \tag{D.2.1}$$

As described in Section 4.2.2, $\mathcal{L}_{ij} = \mathcal{L}_{ij}(p,\sigma)$ can be computed by convolution of the image with the second derivatives of the Gaussian:

$$\mathcal{L}_{ij} = \bar{\mathcal{S}} * \frac{\partial^2}{\partial p_i \partial p_j}\mathcal{G}, \tag{D.2.2}$$

where the 2D Gaussian $\mathcal{G} = \mathcal{G}(p,\sigma)$ at the scale σ is defined as:

$$\mathcal{G}(p,\sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|p\|^2}{2\sigma^2}\right), \tag{D.2.3}$$

where $p = (p_u, p_v)^T$. The second derivatives of the Gaussian are computed as:

$$\frac{\partial^2}{\partial p_i^2}\mathcal{G}(p,\sigma) = \frac{p_i^2 - \sigma^2}{2\pi\sigma^6} \exp\left(-\frac{\|p\|^2}{2\sigma^2}\right), \tag{D.2.4}$$

$$\frac{\partial^2}{\partial p_i \partial p_j}\mathcal{G}(p,\sigma) = \frac{p_i p_j}{2\pi\sigma^6} \exp\left(-\frac{\|p\|^2}{2\sigma^2}\right), \tag{D.2.5}$$

where $i,j \in \{u,v\}$ and $i \neq j$.

Although the Gaussian second derivatives are non-zero for all finite p, they become small fast. In addition, practical implementation requires them to be discretized and truncated. Hence, they have to be expressed as $(2g+1) \times (2g+1)$

[1] Here, we refer to one of the images of the sequence, $\bar{\mathtt{S}}_i$, but the index i is dropped to simplify the notation.

size matrices $\mathtt{G}_{uu}$, $\mathtt{G}_{uv}$ and $\mathtt{G}_{vv}$ (see Figure D.1). The matrix elements are computed as $\mathtt{G}_{ij}[\check{q}] = \frac{\partial^2}{\partial\check{q}_i\partial\check{q}_j}\mathcal{G}(\check{q},\sigma)$, $i,j = \{u,v\}$, where $\check{q} \in \Phi \subset \mathbb{R}^2$, and Φ is a $(2g+1)\times(2g+1)$ grid with the same spacing δ_{px} between the points as in Γ^2, defined as $\Phi = \{(k\delta_{\text{px}}, m\delta_{\text{px}})^T \mid k,m = -g,\ldots,g \in \mathbb{Z}\}$.

Discretization of the Gaussians introduces artifacts in the images resulting from convolution [215]. The larger the g value, the less artifacts are present and the more noise is suppressed, though larger matrices increase convolution time. Based on empirical observations with our data sets, we found that for $\sigma = 0.7$ mm a good noise suppression with the minimal computational effort can be reached by selecting $g = 13$ px, i.e., $\mathtt{G}_{uu}$, $\mathtt{G}_{uv}$, and $\mathtt{G}_{vv}$ are the 27×27 element matrices.

The discrete second derivatives of the scale-space representation $\mathtt{L}_{ij}$ are obtained by a 2D convolution between $\mathtt{G}_{ij}$ and the ROI in the image $\bar{\mathtt{S}}$ ($\bar{\mathtt{S}}_{\text{ROI}}$) defined on $\Gamma^2 \cap \Omega_{\text{ROI}}$. One of the ways to have $\mathtt{L}_{ij}$ of the same size as $\bar{\mathtt{S}}_{\text{ROI}}$ is to extend $\bar{\mathtt{S}}_{\text{ROI}}$ by the margins of g pixels padded with zeros before the convolution. Instead, we fill the margins with the intensity values from $\bar{\mathtt{S}}$, i.e., the image before the convolution, denoted $\bar{\mathtt{S}}_{\text{FL}}$, is formed as a ROI within $\bar{\mathtt{S}}$ whose grid points are in $\Gamma^2 \cap \Omega_{\text{FL}}$ rather than $\Gamma^2 \cap \Omega_{\text{ROI}}$. The Ω_{FL} region is computed from Ω_{ROI} by extending Ω_{ROI} by g pixels in four directions.

The convolution at step $\langle\!\langle \mathbf{F}_{3\text{A}} \rangle\!\rangle$ between $\mathtt{S}_{\text{FL}}$ and $\mathtt{G}_{uu}$, $\mathtt{G}_{uv}$, and $\mathtt{G}_{vv}$ is performed by using Matlab function `conv2` [1]:

$\mathtt{L}_{ij}$ `= conv2(`$\bar{\mathtt{S}}_{\text{FL}}$`,` $\mathtt{G}_{ij}$`, 'valid');`

where $i,j = \{u,v\}$. The Hessian eigenvalues are computed as:

`[·,` D`]` $\leftarrow$ `eig(`H`);`

where D is a matrix with eigenvalues λ_1 and λ_2 on the diagonal, and zeroes elsewhere. After the images $\mathtt{Y}_{\text{m}}$ and $\mathtt{Y}_{\text{mr}}$ are computed at steps $\langle\!\langle \mathbf{F}_{3\text{B}} \rangle\!\rangle$ and $\langle\!\langle \mathbf{F}_{3\text{C}} \rangle\!\rangle$, the dilation $\langle\!\langle \mathbf{F}_{3\text{D}} \rangle\!\rangle$ is performed by first constructing a structuring element `E`:

`E = strel('disk',` $[0.5h/\delta_{\text{px}}]$`, 0);`

where $[0.5h/\delta_{\text{px}}]$ is a radius of the circular structuring element, computed as a half-length of the fiducial marker in pixels, rounded to the closest integer. Then, the dilation of the contrast-enhanced magnitude-and-ratio image $\mathcal{C}(\mathtt{Y}_{\text{mr}})$ is performed by:

$\mathtt{Y}_{\text{d}}$ `= imdilate(`$\mathcal{C}(\mathtt{Y}_{\text{mr}})$`,E);`

Clustering step $\langle\!\langle \mathbf{F}_{3\text{E}} \rangle\!\rangle$ uses a hierarchical clustering technique to combine high-intensity points $\Theta = \{\check{p} \mid \mathtt{Y}_{\text{d}}[\check{p}] \geq \vartheta\}$, into connected clusters. It is invoked as:

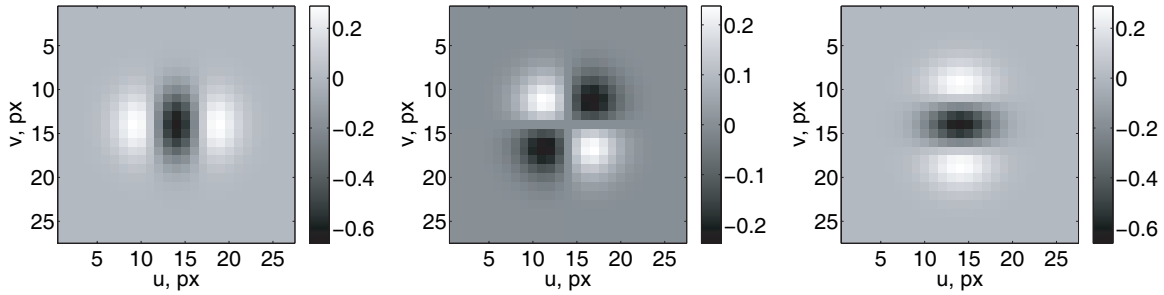


Figure D.1: 27×27 matrices of discrete and truncated Gaussian second derivatives $\mathtt{G}_{uu}$ (*left*), $\mathtt{G}_{uv}$ (*centre*), and $\mathtt{G}_{vv}$ (*right*) for $\sigma = 0.7$ mm.

`T = clusterdata(`$(\check{p})$`, 'cutoff', 1.5`$\,\delta_{\mathrm{px}}$`, 'criterion', 'distance');`

where $(\check{p})$ denotes an array of point coordinates from Θ. The resulting vector `T` contains cluster numbers for the points. Then, all points of each cluster are assigned the maximal intensity value found within the cluster, by which the mask `M` is formed from $\mathtt{Y}_{\mathrm{d}}$. It is applied to image $\mathtt{Y}_{\mathrm{m}}$ by elementwise multiplication of the arrays `M` and $\mathtt{Y}_{\mathrm{m}}$ at step $\langle\!\langle \mathbf{F}_{3\mathrm{F}} \rangle\!\rangle$. Finally, image `R` is computed by application of the contrast enhancement function at step $\langle\!\langle \mathbf{CE}_2 \rangle\!\rangle$.

The MEF implementation is summarized in Algorithm D.6. The details on scale-space representation and computation of the Hessian eigenvalues are given in Algorithm D.7, marker-based contrast enhancement function in Algorithm D.8, and creation of the mask by dilation and flattening in Algorithm D.9.

Algorithm D.4: Temporal filtering.

Input:
$\mathtt{S}_0, \ldots, \mathtt{S}_{m-1}$ – unfiltered fluoroscopic images,
ℓ – filter kernel size

$\mathbf{F}_1$ **Filter coefficients:**
compute ℓ Dolph-Chebyshev coefficients: w
normalize: $w \leftarrow w / \sum_{n=0}^{\ell-1} w_n$

Temporal filtering of images:

$$\bar{\mathtt{S}}_i = \sum_{n=0}^{\ell-1} w_n \, \mathtt{S}_{i+n}, \quad 0 \le i \le m - \ell$$

Output: $\bar{\mathtt{S}}_0, \ldots, \bar{\mathtt{S}}_{m-\ell}$ – filtered fluoroscopic images

Algorithm D.5: Computation of regions of interest.

Input:
$c^k_{3\mathrm{D}}$, $k = 1, \ldots, n$ – marker set/model centres in 3D,
$\mathcal{T}$ – transformation used in 2D-3D registration,
$a^{\max} \in \mathbb{R}^m$ – bounds on the transformation parameters a, where m is a number of degrees of freedom,
$\mathcal{P}$ – 3D to 2D projection operator defined in Eq. (B.5.4),
κ – 3D to 2D scaling factor for the size of the object that is located in the proximity to the point $c^k_{3\mathrm{D}} \in \mathbb{R}^3$ defined in Eq. (B.5.5),
h – marker size

$\mathbf{L_1/T_2}$ **Maximal 2D displacement:**
switch $\mathcal{T}$ **do for** $k \leftarrow 1$ **to** n

case $\mathcal{T}^{2\mathrm{D}}$, *shift only,* $a^{\max} = (t_u^{\max}, t_v^{\max})^T$
$\varrho_k \leftarrow \|a^{\max}\|$

case $\mathcal{T}^{2\mathrm{D}}$, *shift and rotation,* $a^{\max} = (t_u^{\max}, t_v^{\max}, \gamma^{\max})^T$
$\varrho_k \leftarrow \max\limits_a \|\mathcal{P}(c^k_{3\mathrm{D}}) - \mathcal{T}^{2\mathrm{D}}(a, \mathcal{P}(c^k_{3\mathrm{D}}))\|,$
$\quad$ s.t. $-a_j^{\max} \le a_j \le a_j^{\max}$
$\varrho_k \leftarrow \max\{\varrho_k, \|(t_u^{\max}, t_v^{\max})^T\|\}$ % *Safeguard*

case $\mathcal{T}^{3\mathrm{D}}$, *shift only,* $a^{\max} = (t_x^{\max}, t_y^{\max}, t_z^{\max})^T$
$\varrho_k \leftarrow \kappa \|a^{\max}\|$

case $\mathcal{T}^{3\mathrm{D}}$, *shift and rotation,* $a = (t_x, t_y, t_z, \gamma_x, \gamma_y, \gamma_z)^T$
$\varrho_k^{3\mathrm{D}} \leftarrow \max\limits_a \|c^k_{3\mathrm{D}} - \mathcal{T}^{3\mathrm{D}}(a, c^k_{3\mathrm{D}})\|,$
$\quad$ s.t. $-a_j^{\max} \le a_j \le a_j^{\max}$
$\varrho_k \leftarrow \kappa\, \varrho_k^{3\mathrm{D}}$
$\varrho_k \leftarrow \max\{\varrho_k, \|(t_x^{\max}, t_y^{\max}, t_z^{\max})^T\|\}$ % *Safeguard*

end

Extend by half marker size: **for** $k \leftarrow 1$ **to** n **do** $\varrho_k \leftarrow \varrho_k + 0.5\kappa h$

Circular marker ROIs:
for $k \leftarrow 1$ **to** n **do** $\Upsilon_k = \left\{q \in \mathbb{R}^2 \mid \|\mathcal{P}(c^k_{3\mathrm{D}}) - q\| \le \varrho_k\right\}$

Rectangular registration ROI:
$\Omega_{\mathrm{ROI}} = \left\{q \in \mathbb{R}^2 \mid \min\left\{\cup_{k=1}^n \Upsilon_k\right\}_{u,v} \le q_{u,v} \le \max\left\{\cup_{k=1}^n \Upsilon_k\right\}_{u,v}\right\}$

Output: Υ_k and Ω_{ROI} – circular marker and rectangular ROIs

Algorithm D.6: Marker enhancement filter (MEF).

Input:
γ – magnitude parameter,
$\beta = 0.25$ – ratio parameter

Note: All discrete images are processed for $\check{p} \in \Gamma^2 \cap \Omega_{\text{ROI}}$

$\text{F}_{3\text{A}}$ **Hessian eigenvalues of scale-space representation:**
$\lambda_{1,2}[\check{p}]$, Alg. D.7

$\text{F}_{3\text{B}}$ **Magnitude-only image**, Eqs. (4.2.25) and (4.4.34):
$\mathtt{X}_{\text{m}}[\check{p}] = \sqrt{\lambda_1[\check{p}]^2 + \lambda_2[\check{p}]^2}$
$\gamma = 0.5 \max\{\mathtt{X}_{\text{m}}[\check{p}]\}$

if $\check{p} \notin \Gamma^2 \cap \Upsilon_k$, $k = 1, \ldots, n$ **then**
$\mathtt{Y}_{\text{m}}[\check{p}] = 0$
else

$$\mathtt{Y}_{\text{m}}[\check{p}] = \begin{cases} 1 - \exp\left(-\frac{\mathtt{X}_{\text{m}}[\check{p}]^2}{2\gamma^2}\right), & \text{if } \lambda_1[\check{p}] < 0 \text{ and } \lambda_2[\check{p}] < 0, \\ 0, & \text{otherwise} \end{cases}$$

end

$\text{P}_{3\text{C}}$ **Magnitude-and-ratio image**, Eqs. (4.2.26) and (4.4.33):
$\mathtt{X}_{\text{r}}[\check{p}] = \lambda_1[\check{p}]/\lambda_2[\check{p}]$

$$\mathtt{Y}_{\text{mr}}[\check{p}] = \begin{cases} \mathtt{Y}_{\text{m}}[\check{p}] \cdot \left(1 - \exp\left(-\frac{\mathcal{X}_{\text{r}}[\check{p}]^2}{2\beta^2}\right)\right), & \text{if } \lambda_1[\check{p}] < 0 \text{ and } \lambda_2[\check{p}] < 0, \\ 0, & \text{otherwise} \end{cases}$$

CE_1 **Contrast enhancement:**
$\mathcal{C}(\mathtt{Y}_{\text{mr}})$, Alg. D.8

$\text{F}_{3\text{D}-3\text{E}}$ **Mask computation:**
mask $\mathtt{M}$ is produced from $\mathcal{C}(\mathtt{Y}_{\text{mr}})$ by dilation and flattening, Alg. D.9

$\text{F}_{3\text{F}}$ **Mask application to the magnitude-only image:**
$\mathtt{Y}[\check{p}] = \mathtt{M}[\check{p}] \cdot \mathtt{Y}_{\text{m}}[\check{p}]$

CE_2 **Final contrast enhancement:**
$\mathtt{R} = \mathcal{C}(\mathtt{Y})$, Alg. D.8

Output: $\mathtt{R}$ – discrete MEF-image

Algorithm D.7: Scale-space representation and computation of the Hessian eigenvalues.

Input:
$\bar{\mathtt{S}}$ – temporally filtered discrete 2D image,
Ω_{ROI} – rectangular ROI,
σ – scale of the Gaussian,
$g = 13$ px – Gaussian half-size,
δ_{px} – pixel size

F3A **Gaussian grid:**

create grid $\Phi = \{(k\delta_{\text{px}}, m\delta_{\text{px}})^T \mid k, m = -g, \ldots, g \in \mathbb{Z}\}$

Gaussian derivatives:

$$\mathtt{G}_{ii}[\check{q}] = \frac{\partial^2}{\partial \check{q}_i^2}\mathcal{G}(\check{q}, \sigma) = \frac{\check{q}_i^2 - \sigma^2}{2\pi\sigma^6}\,\exp\left(-\frac{\|\check{q}\|^2}{2\sigma^2}\right)$$

$$\mathtt{G}_{ij}[\check{q}] = \frac{\partial^2}{\partial \check{q}_i \partial \check{q}_j}\mathcal{G}(\check{q}, \sigma) = \frac{\check{q}_i \check{q}_j}{2\pi\sigma^6}\,\exp\left(-\frac{\|\check{q}\|^2}{2\sigma^2}\right)$$

where $\check{q} = (\check{q}_u, \check{q}_v)^T \in \Phi$, $i, j \in \{u, v\}$ and $i \neq j$

Filtering ROI:

$\Omega_{\text{FL}} = \{p \in \mathbb{R}^2 \mid \min(\Omega_{\text{ROI}})_{u,v} - g\delta_{\text{px}} \leq p_{u,v} \leq \max(\Omega_{\text{ROI}})_{u,v} + g\delta_{\text{px}}\}$

$\mathtt{S}_{\text{FL}}[\check{p}] = \bar{\mathtt{S}}[\check{p}]$, where $\check{p} \in \Gamma^2 \cap \Omega_{\text{FL}}$

Derivatives of scale-space representation:

$\mathtt{L}_{ij} =$ convolve $\mathtt{S}_{\text{FL}}$ and $\mathtt{G}_{ij}$, cut to $\Gamma^2 \cap \Omega_{\text{ROI}}$

Hessian matrices: $H[\check{p}] = \begin{bmatrix} \mathtt{L}_{uu}[\check{p}] & \mathtt{L}_{uv}[\check{p}] \\ \mathtt{L}_{uv}[\check{p}] & \mathtt{L}_{vv}[\check{p}] \end{bmatrix}$

Eigenvalues of the Hessians:

for $\check{p} \in \Gamma^2 \cap \Omega_{\text{ROI}}$ **do**
  find D, diagonal matrix with $H[\check{p}]$ eigenvalues
  if $|D_{11}| \leq |D_{22}|$ **then**
    $\lambda_1[\check{p}] = D_{11}$, $\lambda_2[\check{p}] = D_{22}$
  else
    $\lambda_1[\check{p}] = D_{22}$, $\lambda_2[\check{p}] = D_{11}$
  end
end

Output: $\lambda_{1,2}[\check{p}]$, $\check{p} \in \Gamma^2 \cap \Omega_{\text{ROI}}$ – Hessian eigenvalues of scale-space representation for all points of the image

Algorithm D.8: Marker-based smooth contrast enhancement.

Input:
$\mathtt{Y}$ – discrete image,
n – number of fiducial markers,
r, h – radius and length of the marker,
δ_{px} – pixel size,
$\tau = 0.5$ – minimal intensity of the candidate marker points,
$s = 3$ – sharpness of distinction between marker and non-marker points

CE **Marker-based threshold:**
number of marker points in the image: $m = n\lceil 2rh/\delta_{\text{px}}^2 \rceil$
sort $\mathtt{Y}[\check{p}]$ in a descending order: $y_1 \geq y_2 \geq \dots$
$\vartheta = y_m$

Contrast enhancement, Eqs. (4.4.37) and (4.4.40):
$$\mathcal{Z}(\mathtt{Y}[\check{p}]) = \frac{\vartheta^s \log_\vartheta \tau}{\vartheta^s - 1}\left(1 - \frac{1}{(\mathtt{Y}[\check{p}])^s}\right)$$
$$\mathcal{C}(\mathtt{Y}[\check{p}]) = \mathtt{Y}[\check{p}]^{\mathcal{Z}(\mathtt{Y}[\check{p}])}$$

Output: $\mathcal{C}(\mathtt{Y})$ – contrast-enhanced image, ϑ – marker-based intensity threshold

Algorithm D.9: Creation of the mask image by dilation and flattening.

Input:
h – length of the fiducial marker,
δ_{px} – pixel size,
$\mathcal{C}(\mathtt{Y}_{\text{mr}})$ – magnitude-and-ratio image after contrast enhancement,
ϑ – marker-based intensity threshold computed in Algorithm D.8

F3D **Dilation:**
create a structuring element $\mathtt{E}$, a circle of $[0.5h/\delta_{\text{px}}]$ px radius
$\mathtt{Y}_{\text{d}} \leftarrow$ dilate $\mathcal{C}(\mathtt{Y}_{\text{mr}})$ with $\mathtt{E}$

F3E **Flattening:**
$\mathtt{M} \leftarrow \mathtt{Y}_{\text{d}}$
points enhanced at $\langle\!\langle \mathbf{CE}_1 \rangle\!\rangle$ step: $\Theta = \{\check{p} \mid \mathtt{Y}_{\text{d}}[\check{p}] \geq \vartheta\}$
separate into connected subsets: $[\mathrm{T}_1, \dots, \mathrm{T}_m] \leftarrow \text{cluster}(\Theta, \eta = 1.5\,\delta_{\text{px}})$
by hierarchical clustering described in step $\langle\!\langle \mathbf{S}_2 \rangle\!\rangle$ of Alg. D.1
for $k \leftarrow 1$ **to** m **do**
$v \leftarrow \max\{\mathtt{Y}_{\text{d}}[\check{p}] \mid \check{p} \in \mathrm{T}_k\}$
$\mathtt{M}[\check{p} \in \mathrm{T}_k] \leftarrow v$
end

Output: $\mathtt{M}$ – mask image

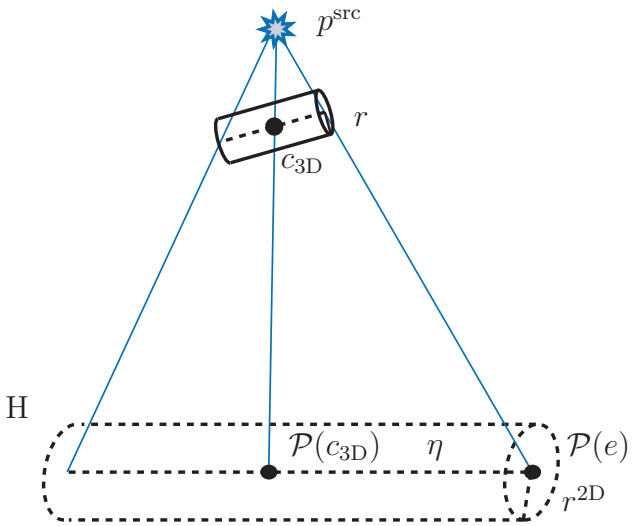


Figure D.2: Extent of projection of the marker model.

D.3 2D Template Image Generation

Prior to discussing the implementation details and providing algorithm listings in Appendix D.3.2, we define the regions of points in the generated template, for which the computation should be performed in Appendix D.3.1. Those are the points that can potentially contain the "projections" of the 3D markers, and can be computed based on the positions (and orientations) of the markers in 3D.

D.3.1 Extents of Generated Markers

Template images consist of the generated bright markers on a flat dark background. For computational efficiency, a set of pixels, for which the DRR computation is performed, or the extents of the generated markers, is determined at the preprocessing step for any of the three DRR generation methods. For the marker model DRR (MM DRR), the extent is comprised of the pixels that can potentially contain a non-zero intensity, i.e., the rays projected from the x-ray source to the centres of these pixels have the potential to intersect the 3D marker model, M^{mod}. For the voxel splat and marker splat DRRs (VS and MS DRRs), the extent is determined as the smallest region outside of which the corresponding Gaussians can be truncated without unreasonable sacrifice of image quality as will be discussed later in this section.

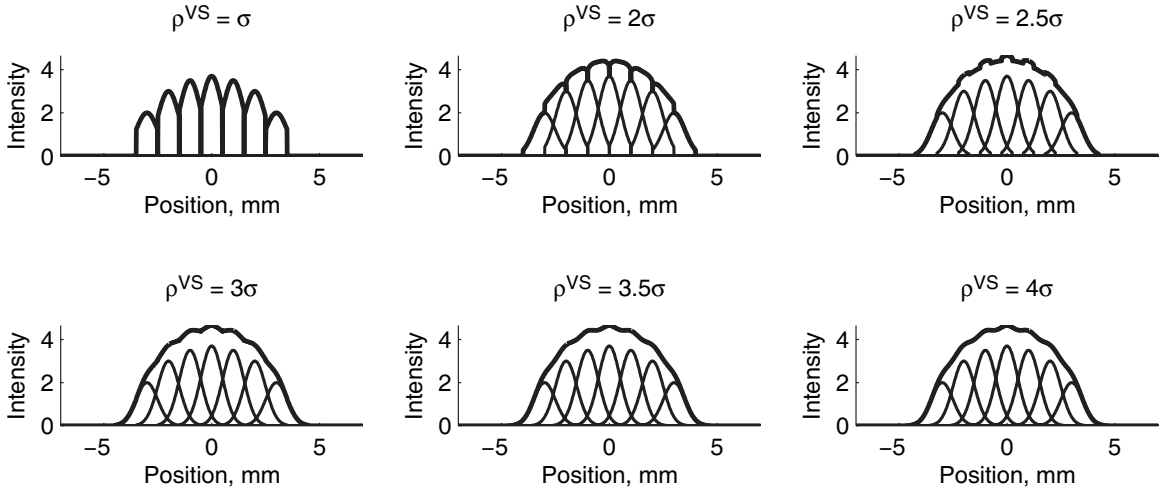


Figure D.3: The use of the voxel splat Gaussians (*thin lines*) computed over small extents ρ_i^{VS} also affects the smoothness of their sum (*bold lines*). We select $\rho_i^{\text{VS}} = 3\sigma_i$.

▲ ρ^{MM}, **Extents of the marker model projections**

We define a set of points on the isocentric plane H, which belong to the rays that can potentially go through the cylindrical marker model, as a circular region around the projection of the marker model centre $c_{3\text{D}}$:

$$\Phi^{\text{MM}} = \{q \in \mathbb{R}^2 \mid \|q - \mathcal{P}(c_{3\text{D}})\| \leq \rho^{\text{MM}}\}, \tag{D.3.1}$$

where $\mathcal{P}$ is the 3D to 2D projection operator defined in Equation (B.5.4), and ρ^{MM} is the maximal dimension of the projection of the marker model from the point $\mathcal{P}(c_{3\text{D}})$ computed as shown in Figure D.2 and explained below. Let e^1, e^2 be the endpoints of the k^{th} cylindrical marker model and r its radius. Let

$$\eta = \max\left\{\|\mathcal{P}(c_{3\text{D}}) - \mathcal{P}(e^1)\|, \|\mathcal{P}(c_{3\text{D}}) - \mathcal{P}(e^2)\|\right\}. \tag{D.3.2}$$

Then, ρ^{MM} can be computed as $\eta + r^{2\text{D}}$, where $r^{2\text{D}} = \alpha\, r$, and α is the 3D to 2D scaling factor for the size of the object that is located in the proximity to the point $c_{3\text{D}}$ defined in Equation (B.5.5). Similar approach can be undertaken to compute the extent of the circular markers. ■

▲ ρ_i^{VS}, **Extents of the voxel splats**

Similar to the MM DRR, the values of each voxel splat should only be computed at the points that are close to the points $\mathcal{P}(\check{p}^i)$, where $\check{p}^i$ are the centres of the segmented

CBCT voxels, as the Gaussians used for the voxel splats become small fast. We define the voxel splat extent as $\Phi_i^{\text{VS}} = \{q \in \mathbb{R}^2 \mid \|q - \mathcal{P}(\check{p}^i)\| \leq \rho_i^{\text{VS}}\}$, where $\rho_i^{\text{VS}} = 3\sigma_i$ and σ_i are the voxel splat Gaussian scales defined in Section 5.1.2. As can be seen from the 1D example in Figure D.3, the use of shorter extents introduces noticeable and undesirable intensity variations in the sum of the splats (*bold lines*), while the use of the splats with $\rho_i^{\text{VS}} > 3\sigma_i$ increases computational cost but does not improve the smoothness of the generated markers significantly. ■

▲ ρ^{MS}, **Extents of the marker splats**

For efficiency, the Gaussian values for the marker splats are only computed for the points that lie in the proximity to the projected marker centres: $\Phi^{\text{MS}} = \{q \in \mathbb{R}^2 \mid \|q - \mathcal{P}(c_{\text{3D}})\| \leq \rho^{\text{MS}}\}$. Similar to the considerations on the voxel splat extents explained above, we select the extents of the marker splats to be $\rho^{\text{MS}} = 3\sigma$, where σ is the scale of the marker splat Gaussian defined in Section 5.1.3. ■

D.3.2 Implementation and Algorithms

In this section, we provide implementation details and algorithm listings for template image generation step $\langle\!\langle \mathbf{L}_{\text{3A}} \rangle\!\rangle$ of the marker localization procedure described in Section 5.1. Algorithms D.10 – D.12 summarize the MM DRR generation method for cylindrical and spherical markers described in Section 5.1.1. The detailed formulations of the marker models are given in Appendix C.3. The VS and MS DRR methods described in Sections 5.1.2 and 5.1.3, respectively, are summarized in Algorithms D.13 and D.14.

Common to all three methods, the discrete 2D template image `W` is initialized with zeros in all points $\check{q} \in \Gamma^2 \cap \Omega_{\text{LOC}}$, where Ω_{LOC} is the registration ROI computed as explained in Section 4.3 and Algorithm D.5. Then, for each marker model, voxel or marker splat, a corresponding region of interest within $\Gamma^2 \cap \Omega_{\text{LOC}}$ is computed, denoted by $\check{\Phi}^{\text{MM}}$, $\check{\Phi}^{\text{VS}}$, and $\check{\Phi}^{\text{MS}}$, respectively. This is done to reduce computational cost by only evaluating the template image intensity values in small relevant parts of the image. Furthermore, for the VS and MS DRRs, we perform computation on rectangular regions that include $\check{\Phi}^{\text{VS}}$ and $\check{\Phi}^{\text{MS}}$, as it simplifies implementation and, in fact, reduces computational times. After the markers are generated, the image `W` is rescaled from 0 to 1 to occupy an intensity range similar to that of the MEF-image.

Algorithm D.10: Template image generation for cylindrical marker models.

Input:
Γ^2 – pixel grid,
$\Omega_{\text{LOC}} \subset \mathbb{R}^2$ – registration region of interest,
$p^{\text{src}} \in \mathbb{R}^3$ – x-ray source location,
r, $c^k_{\text{3D}}, e^{1,k}, e^{2,k} \in \mathbb{R}^3$, $k = 1, \ldots, n$ – marker model radius, centres, endpoints,
ϕ – angle of rotation of the imaging system around the Y axis,
χ_{SAD} – distance between the x-ray source p^{src} and the isocentric plane H

L3A **Initialize:** $p^{\text{src}} \leftarrow R_y(-\phi)\, p^{\text{src}}$

Fill in default background: $\mathtt{W}[\check{q}] \leftarrow 0$, $\check{q} \in \Gamma^2 \cap \Omega_{\text{LOC}}$

for $k \leftarrow 1$ **to** n **do**

Rotate to the initial position:
$c^k_{\text{3D}} \leftarrow R_y(-\phi)\, c^k_{\text{3D}}$, $e^{1,k} \leftarrow R_y(-\phi)\, e^{1,k}$, $e^{2,k} \leftarrow R_y(-\phi)\, e^{2,k}$

Compute A^{cyl}, Alg. D.11

Compute $\check{\Phi}^{\text{MM}}$**:** $r^{\text{2D}} \leftarrow \mathtt{Scale3DTo2D}\ (r,\, c^k_{\text{3D}},\, \phi,\, \chi_{\text{SAD}})$, Alg. D.15
$\rho^{\text{MM}} \leftarrow r^{\text{2D}} + \max\left\{\|\mathcal{P}(c^k_{\text{3D}}) - \mathcal{P}(e^{1,k})\|, \|\mathcal{P}(c^k_{\text{3D}}) - \mathcal{P}(e^{2,k})\|\right\}$
$\check{\Phi}^{\text{MM}} \leftarrow \Gamma^2 \cap \left\{q \in \mathbb{R}^2 \mid \|q - \mathcal{P}(c^k_{\text{3D}})\| \le \rho^{\text{MM}}\right\}$

for each $\check{q} \in \check{\Phi}^{\text{MM}}$ **do**

$p^{\text{iso}} \leftarrow (0, -\check{q}_v, -\check{q}_u, 1)^T$

Solve for $s^{a,b}$ **(and** $\lambda_{a,b}$**):** $\begin{cases} s^T A^{\text{cyl}}\, s = 0, \\ s = p^{\text{src}} + \lambda\,(p^{\text{iso}} - p^{\text{src}}) \end{cases}$

if s^a and s^b are real **then**

if s^a and s^b satisfy (5.1.2) **then**
$\mathtt{W}[\check{q}] \leftarrow \mathtt{W}[\check{q}] + \|s^a - s^b\|$

else

Solve for s^c **(and** λ_c**):** $\begin{cases} (s - p^k_{\text{e}_1})^T (p^k_{\text{e}_1} - p^k_{\text{c}}) = 0, \\ s = p^{\text{src}} + \lambda\,(p^{\text{iso}} - p^{\text{src}}) \end{cases}$

Solve for s^d **(and** λ_d**):** $\begin{cases} (s - p^k_{\text{e}_2})^T (p^k_{\text{e}_2} - p^k_{\text{c}}) = 0, \\ s = p^{\text{src}} + \lambda\,(p^{\text{iso}} - p^{\text{src}}) \end{cases}$

Select $s^{\ell_1}, \ldots, s^{\ell_m}$ from $s^a, \ldots, s^d$ that satisfy (5.1.2)

$\mathtt{W}[\check{q}] \leftarrow \mathtt{W}[\check{q}] + \max\left\{\|s^{\ell_i} - s^{\ell_j}\|,\ i, j = 1, \ldots, m\right\}$

end

end

end

end

Rescale $\mathtt{W}$ to $[0, 1]$ range

Output: $\mathtt{W}$ – marker model DRR for cylindrical markers defined on $\Gamma^2 \cap \Omega_{\text{LOC}}$

Algorithm D.11: Cylinder model transformation.

Input:
r, h – marker model radius and length,
$c_{3\mathrm{D}}$, e – centre and one of the endpoints of the marker model (in homogeneous coordinates)

Cylindrical shape:

$$Q \leftarrow \begin{pmatrix} 1/r^2 & 0 & 0 & 0 \\ 0 & 1/r^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Geometric transformation:

$p^{\mathrm{sh}} \leftarrow e - c_{3\mathrm{D}}$

$d \leftarrow \sqrt{(p_x^{\mathrm{sh}})^2 + (p_y^{\mathrm{sh}})^2}$

if $d > 0$ **then**
% *Marker model axis is not parallel to Z*

$$\sin\gamma_z \leftarrow \frac{p_x^{\mathrm{sh}}}{d}, \quad \cos\gamma_z \leftarrow \frac{p_y^{\mathrm{sh}}}{d}, \quad R_z(-\gamma_z) \leftarrow \begin{pmatrix} \cos\gamma_z & -\sin\gamma_z & 0 & 0 \\ \sin\gamma_z & \cos\gamma_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$p^{\mathrm{rot}} \leftarrow R_z(-\gamma_z)\, p^{\mathrm{sh}}$

$$\sin\gamma_x \leftarrow \frac{p_y^{\mathrm{rot}}}{h/2}, \quad \cos\gamma_x \leftarrow \frac{p_z^{\mathrm{rot}}}{h/2}, \quad R_x(-\gamma_x) \leftarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\gamma_x & -\sin\gamma_x & 0 \\ 0 & \sin\gamma_x & \cos\gamma_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$F \leftarrow R_x(-\gamma_x) R_z(-\gamma_z) T^{3\mathrm{D}}(-c_{3\mathrm{D}})$

else
$F \leftarrow T^{3\mathrm{D}}(-c_{3\mathrm{D}})$

end

Transformed cylindrical shape: $A^{\mathrm{cyl}} \leftarrow F^T Q F$

Output: A^{cyl} – matrix of the transformed cylindrical shape

Algorithm D.12: Template image generation for spherical marker models.

Input:
Γ^2 – pixel grid,
$\Omega_{\text{LOC}} \subset \mathbb{R}^2$ – registration region of interest,
$p^{\text{src}} \in \mathbb{R}^3$ – x-ray source location,
r_k, $c^k_{\text{3D}} \in \mathbb{R}^3$, $k = 1, \ldots, n$ – radii and centres of the marker models
ϕ – angle of rotation of the imaging system around the Y axis,
χ_{SAD} – distance between the x-ray source p^{src} and the isocentric plane H

$\text{L}_{3\text{A}}$ **Initialize:** $p^{\text{src}} \leftarrow R_y(-\phi)\, p^{\text{src}}$

Fill in default background: $\mathtt{W}[\check{q}] \leftarrow 0$, $\check{q} \in \Gamma^2 \cap \Omega_{\text{LOC}}$

for $k \leftarrow 1$ **to** n **do**

Rotate to the initial position: $c^k_{\text{3D}} \leftarrow R_y(-\phi)\, c^k_{\text{3D}}$

Compute $\check{\Phi}^{\text{MM}}$**:**

$\rho^{\text{MM}} \leftarrow$ `Scale3DTo2D` $(r, c^k_{\text{3D}}, \phi, \chi_{\text{SAD}})$, Alg. D.15

$\check{\Phi}^{\text{MM}} \leftarrow \Gamma^2 \cap \left\{ q \in \mathbb{R}^2 \mid \|q - \mathcal{P}(c^k_{\text{3D}})\| \le \rho^{\text{MM}} \right\}$

Compute A^{sph} as defined in Eq. (C.3.9)

for each $\check{q} \in \check{\Phi}^{\text{MM}}$ **do**

$p^{\text{iso}} \leftarrow (0, -\check{q}_v, -\check{q}_u, 1)^T$

Solve for $s^{a,b}$ **(and** $\lambda_{a,b}$**):** $\begin{cases} s^T A^{\text{sph}}\, s = 0, \\ s = p^{\text{src}} + \lambda\,(p^{\text{iso}} - p^{\text{src}}) \end{cases}$

if s^a and s^b are real **then**

$\mathtt{W}[\check{q}] \leftarrow \mathtt{W}[\check{q}] + \|s^a - s^b\|$

end

end

end

Rescale `W` to $[0, 1]$ range

Output: `W` – marker model DRR for spherical markers defined on $\Gamma^2 \cap \Omega_{\text{LOC}}$

Algorithm D.13: Voxel spat DRR generation.

Input:
Γ^2 – pixel grid,
$\Omega_{\text{LOC}} \subset \mathbb{R}^2$ – registration region of interest,
$c^k_{\text{3D}} \in \mathbb{R}^3$, $k = 1, \ldots, n$ – marker model centres,
δ_{vx} – voxel size,
ϕ – angle of rotation of the imaging system around the Y axis,
χ_{SAD} – distance between the x-ray source p^{src} and the isocentric plane H

L3A **Initialize:** Fill in default background: $\mathtt{W}[\breve{q}] \leftarrow 0$, $\breve{q} \in \Gamma^2 \cap \Omega_{\text{LOC}}$

for $k \leftarrow 1$ **to** n **do**
  for $i \leftarrow 1$ **to** $|\text{M}_k|$ **do**
    Compute $\breve{\Phi}^{\text{VS}}$ **for** $p^i \in \text{M}_k$**:**
    $\delta^{\text{2D}}_{\text{vx}} \leftarrow$ `Scale3DTo2D` $(\delta_{\text{vx}}, p^i, \phi, \chi_{\text{SAD}})$, Alg. D.15
    $\sigma^i \leftarrow 0.5\, \delta^{\text{2D}}_{\text{vx}}$
    $\rho^{\text{VS}} \leftarrow 3\, \sigma^i$
    $\breve{\Phi}^{\text{VS}} \leftarrow \Gamma^2 \cap \left\{ q \in \mathbb{R}^2 \mid \mathcal{P}(p^i)_{u,v} - \rho^{\text{VS}} \leq q_{u,v} \leq \mathcal{P}(p^i)_{u,v} + \rho^{\text{VS}} \right\}$
    for each $\breve{q} \in \breve{\Phi}^{\text{VS}}$ **do**
      $\mathtt{W}[\breve{q}] \leftarrow \mathtt{W}[\breve{q}] + \mathcal{G}(\breve{q} - \mathcal{P}(p^i), \sigma^i)$
    end
  end
end

Rescale `W` to $[0, 1]$ range

Output: `W` – voxel splat DRR defined on $\Gamma^2 \cap \Omega_{\text{LOC}}$

Algorithm D.14: Marker splat DRR generation.

Input:
Γ^2 – pixel grid,
$\Omega_{\text{LOC}} \subset \mathbb{R}^2$ – registration region of interest,
$c^k_{\text{3D}} \in \mathbb{R}^3$, $k = 1, \ldots, n$ – marker model centres,
$\tilde{\sigma}$ – Gaussian scale for the marker splat, Sect. 5.1.3,
ϕ – angle of rotation of the imaging system around the Y axis,
χ_{SAD} – distance between the x-ray source p^{src} and the isocentric plane H

L3A **Initialize:** Fill in default background: $\texttt{W}[\check{q}] \leftarrow 0$, $\check{q} \in \Gamma^2 \cap \Omega_{\text{LOC}}$

for $k \leftarrow 1$ **to** n **do**

Compute $\check{\Phi}^{\text{MS}}$**:**

$\sigma \leftarrow$ `Scale3DTo2D` $(\tilde{\sigma}, c^k_{\text{3D}}, \phi, \chi_{\text{SAD}})$, Alg. D.15

$\rho^{\text{MS}} \leftarrow 3\,\sigma$

$\check{\Phi}^{\text{MS}} \leftarrow \Gamma^2 \cap \left\{ q \in \mathbb{R}^2 \mid \mathcal{P}(c^k_{\text{3D}})_{u,v} - \rho^{\text{MS}} \leq q_{u,v} \leq \mathcal{P}(c^k_{\text{3D}})_{u,v} + \rho^{\text{MS}} \right\}$

for each $\check{q} \in \check{\Phi}^{\text{MS}}$ **do**

$\texttt{W}[\check{q}] \leftarrow \texttt{W}[\check{q}] + \mathcal{G}(\check{q} - \mathcal{P}(c^k_{\text{3D}}), \sigma)$

end

end

Rescale `W` to $[0, 1]$ range

Output: `W` – marker position DRR defined on $\Gamma^2 \cap \Omega_{\text{LOC}}$

Algorithm D.15: Scaling a size of the object from 3D to 2D.

Function $\delta^{\text{2D}} =$ `Scale3DTo2D` $(\delta^{\text{3D}}, p, \phi, \chi_{\text{SAD}})$

Input:
δ^{3D} – length of the object in 3D,
p – point in 3D in the proximity or within the object,
ϕ – angle of rotation of the imaging system around the Y axis,
χ_{SAD} – distance between the x-ray source p^{src} and the isocentric plane H

Scale from 3D to 2D:

$\kappa \leftarrow \dfrac{\chi_{\text{SAD}}}{\chi_{\text{SAD}} - (R_y(-\phi)\, p)_x}$, Eq. (B.5.6)

$\delta^{\text{2D}} \leftarrow \kappa\, \delta^{\text{3D}}$, Eq. (B.5.5)

Output: δ^{2D} – length scaled to 2D

D.4 2D Marker Localization

D.4.1 Optimization Starting Point

As explained in Section 5.3.3, in order to provide a reasonable starting point for the optimization procedure that is used in image registration between the generated template and the MEF-image, we apply the normalized 2D cross-correlation method with multiple templates (NCC-MT) at reduced resolution. Several template images at different rotation angles $\{\beta_i\}$ are generated, and downsampled, together with the MEF-image, by applying a 3×3 pixel averaging. Then, the NCC is computed by the Matlab function `normxcorr2` between the MEF-image and each of the templates, producing matrices $C_i[du, dv]$, where du and dv correspond to the shifts of the i^{th} template relative to the MEF-image in the U and V directions, respectively. Let $C_m[u, v]$ be the largest correlation value among all values in all C_i matrices. Then, the rotation angle of the optimization starting point, $a^0 = (t_u^0, t_v^0, \gamma^0)^T$, equals to the template rotation angle, $\gamma^0 = \beta_m$. However, the found shifts u and v do not directly correspond to t_u^0 and t_v^0 due to the fact that the NCC-MT approach assumes the application of rotation before the shifts, while the geometric transformations defined in this thesis, and used in our marker localization and tracking procedures, assume otherwise (see Appendix B). Given that the 2D geometric transformation is used in registration, the elements of the optimization starting point a^0 can be computed from the NCC shifts, u and v, and the rotation angle, β_m, as follows:

$$\begin{aligned} t_u^0 &= \quad u\cos\beta_m + v\sin\beta_m, \\ t_v^0 &= -u\sin\beta_m + v\cos\beta_m, \\ \gamma^0 &= \quad \beta_m. \end{aligned} \tag{D.4.1}$$

D.4.2 Implementation and Algorithms

This section provides implementation details for the 2D marker localization procedure described in Chapter 5. The approach is summarized in Algorithm D.16. We start by estimating the calibration parameters for the fluoroscopic frames from the corresponding parameters of the CBCT projections (Appendix B.4). This is necessary to establish the correspondence between the $\tilde{\text{U}}\tilde{\text{V}}$ (imaging plane $\tilde{\text{H}}$, pixels) and UV (isocentric plane H, millimeters) coordinate systems, and compute the coordinates of

the grids points in millimeters. Then, the marker model enlargement is performed as explained in Section 5.1.1 if the MM DRR template image generation method is used.[2]

The algorithm proceeds with the computation of the localization ROI, Ω_{LOC}, referred to as step $\langle\!\langle \mathbf{L}_1 \rangle\!\rangle$ in Flowchart 5.1 and described earlier in Algorithm D.5. At step $\langle\!\langle \mathbf{L}_2 \rangle\!\rangle$, the control is passed to the filtering block $\langle\!\langle \mathbf{F} \rangle\!\rangle$ that processes fluoroscopic images to obtain the MEF-image `R`. Then, at step $\langle\!\langle \mathbf{L}_3 \rangle\!\rangle$, `R` is registered to the template `W` computed as explained in Appendix D.3. Optimal geometric parameters, a^*, that bring into alignment `R` and the transformed template, $\mathtt{W}_a$, are computed by solving (5.3.10) using SQP implemented in Matlab function `fmincon` (active set) that iteratively minimizes $\mathcal{D}_{\text{NNCC}}$, defined in Section C.4, over Ω_{LOC}. The optimization search starts with the point a^0 that is computed as explained in Section 5.3.3 and Appendix D.4.1. At step $\langle\!\langle \mathbf{L}_4 \rangle\!\rangle$, the 2D marker positions c^k_{2D} are computed by applying a geometric transformation with the optimal parameters a^* found at step $\langle\!\langle \mathbf{L}_3 \rangle\!\rangle$ to the 3D marker positions, c^k_{3D}. Note that in Algorithm D.16 we will use the notation $c^k_{\text{2D}} = \mathcal{T}(a^*, c^k_{\text{3D}})$ to denote an operator that transforms the points c^k_{3D} by either the 3D or 2D rigid geometric transformations, $\mathcal{T}^{\text{3D}}$ or $\mathcal{T}^{\text{2D}}$, respectively:

$$c^k_{\text{2D}} = \mathcal{T}(a^*, c^k_{\text{3D}}) = \mathcal{P}(\mathcal{T}^{\text{3D}}(a^*, c^k_{\text{3D}})), \tag{D.4.2}$$

or

$$c^k_{\text{2D}} = \mathcal{T}(a^*, c^k_{\text{3D}}) = \mathcal{T}^{\text{2D}}(a^*, \mathcal{P}(c^k_{\text{3D}})) \tag{D.4.3}$$

where $\mathcal{T}^{\text{3D}}$ and $\mathcal{T}^{\text{2D}}$, and their corresponding parameter vectors a, are defined in Appendices B.1 and B.2, respectively, and $\mathcal{P}$ is the 3D to 2D projection operator defined in Equation (B.5.4).

[2] Although we decided to use the MS DRR method to process our images (Section 5.1), the description provided here corresponds to the Gryphon implementation, which allows choosing different modelling options, including image generation methods and dimensionality of the geometric transformation.

Algorithm D.16: Marker localization in 2D space.

Input:
M_k or $\mathrm{M}_k^{\mathrm{mod}}$ – 3D marker sets or marker models,
$c_{3\mathrm{D}}^k \in \mathbb{R}^3$ – 3D marker positions

Initialize:
Estimate calibration parameters for fluoroscopic frames,
express grid point coordinates in millimeters instead of pixels, App. B.4

if MM DRR used **then**
  Enlarge $\mathrm{M}_k^{\mathrm{mod}}$, Sect. 5.1.1
end

L1 **Compute localization ROI:** Ω_{LOC}, using $a^{\max}$ in Alg. D.5

L2 **Open and filter fluoroscopic images:** block $\langle\!\langle \mathbf{F} \rangle\!\rangle$, obtain MEF-image $\mathtt{R}$

L3 **Register images:**

Define optimization starting point a^0, App. D.4.1

L3B Call `fmincon` to solve (5.3.10):

$$a^* = \arg\min_a \mathcal{D}_{\mathrm{NCC}}\left(\mathtt{W}_a, \mathtt{R}\right), \quad \text{s.t. } |a_j| \le a_j^{\max},$$

where:

L3A $\mathtt{W}_a$ is computed in App. D.3, Algs. D.10 – D.14,
$\mathcal{D}_{\mathrm{NNCC}}$ defined in Sect. 5.2 and App. C.4,
and $a^{\max}$ in Sect. 5.3.1

L4 **Compute** $c_{2\mathrm{D}}^k = \mathcal{T}(a^*, c_{3\mathrm{D}}^k)$

Output: a^* – optimal transformation parameters, $c_{2\mathrm{D}}^k \in \mathbb{R}^2$ – 2D marker positions

D.5 Marker Motion Tracking

D.5.1 Optimization Starting Point

As explained in Section 6.3, the optimization starting point $a^0 = (t_u^0, t_v^0, \gamma^0)^T$ is computed as the combination of the prediction $\hat{a} = (\hat{t}_u, \hat{t}_v, \hat{\gamma})^T$ obtained by using the linear regression model based on three preceding images of the sequence (LR-3), and the point $(u, v, \beta)^T$, where the normalized cross-correlation (NCC) between the incoming image and the template transformed by $\hat{a}$ reach its maximum. In contrast to 2D-3D image registration used in the marker localization procedure (Section 5.3 and Appendix D.4.1), the displacements between the markers in the neighbouring images of the sequence are expected to be smaller than between the CBCT and the fluoroscopic sequence. Therefore, we do not use multiple templates with the NCC (multiple rotation angles), i.e., $\beta = 0°$. The elements of the starting point a^0 are computed as:

$$\begin{aligned} t_u^0 &= \hat{t}_u + u, \\ t_v^0 &= \hat{t}_v + v, \\ \gamma^0 &= \hat{\gamma}. \end{aligned} \tag{D.5.1}$$

D.5.2 Implementation and Algorithms

This section provides implementation details for the marker tracking procedure described in Chapter 6. The approach is summarized in Algorithm D.18. The details on the 2D template image generation method used in 2D-2D image registration can be found in Algorithm D.17. For clarity, Algorithm D.18 describes tracking in one image sequence. The Gryphon software, however, is built in a way that allows creating a list of sequences to process (the same is done for the 2D marker localization procedure described in Appendix D.4).

We start by defining the ROI Ω_{LOC}, which is the same as in 2D marker localization. This ROI is used to perform the initial reduction of the image size. Registration in tracking is performed on Ω_{TR} (computed at step $\langle\langle \mathbf{T}_2 \rangle\rangle$), that is different for every image of the sequence and depends on $\hat{a}$ (a prediction for geometric transformation parameters computed at step $\langle\langle \mathbf{T}_1 \rangle\rangle$) and $b^{\max}$ (bounds on displacements between the markers in consecutive images defined in Section 6.1).

After the image is opened and preprocessed by temporal and marker enhancement filtering at step $\langle\!\langle \mathbf{T}_3 \rangle\!\rangle$, it can be registered to the template, generated by the modification of the marker splat DRR method (Section 5.1.3, Algorithm D.14) for the 2D (Section 6.1, Algorithm D.17) at step $\langle\!\langle \mathbf{T}_4 \rangle\!\rangle$. Finally, the resulting geometric transformation parameters $(a^*)^i$ and the corresponding 2D marker positions for image i, $(c^k_{2\mathrm{D}})^i$, are saved at step $\langle\!\langle \mathbf{T}_5 \rangle\!\rangle$. After the procedure described above is performed for each image of the sequence, the final results, in particular, the displacements $(d^k)^i$ of the markers from their positions in the first image of the sequence, $(c^k_{2\mathrm{D}})^0$, are produced at step $\langle\!\langle \mathbf{T}_6 \rangle\!\rangle$.

Algorithm D.17: 2D marker splat template image generation.

Input:
Γ^2 – pixel grid,
$\Omega_{\mathrm{LOC}} \subset \mathbb{R}^2$ – region of interest,
$c^k_{2\mathrm{D}} \in \mathbb{R}^2$, $k = 1, \ldots, n$ – marker centres,
$\tilde{\sigma}$ – Gaussian scale for the marker splat, Sect. 5.1.3

$\mathbf{T}_{4\mathrm{A}}$ **Initialize:** Fill in default background: $\mathtt{W}[\check{q}] \leftarrow 0$, $\check{q} \in \Gamma^2 \cap \Omega_{\mathrm{LOC}}$

for $k \leftarrow 1$ **to** n **do**

Compute $\check{\Phi}^{\mathrm{MS}}$**:**

$\rho^{\mathrm{MS}} \leftarrow 3\,\tilde{\sigma}$

$\check{\Phi}^{\mathrm{MS}} \leftarrow \Gamma^2 \cap \left\{ q \in \mathbb{R}^2 \mid (c^k_{2\mathrm{D}})_{u,v} - \rho^{\mathrm{MS}} \leq q_{u,v} \leq (c^k_{2\mathrm{D}})_{u,v} + \rho^{\mathrm{MS}} \right\}$

for each $\check{q} \in \check{\Phi}^{\mathrm{MS}}$ **do**

$\mathtt{W}[\check{q}] \leftarrow \mathtt{W}[\check{q}] + \mathcal{G}(\check{q} - c^k_{2\mathrm{D}}, \sigma)$

end

end

Rescale $\mathtt{W}$ to $[0, 1]$ range

Output: $\mathtt{W}$ – 2D marker splat template image defined on $\Gamma^2 \cap \Omega_{\mathrm{LOC}}$

Algorithm D.18: Marker tracking.

Input:
$(c^k_{2\mathrm{D}})^0 \in \mathbb{R}^2$ – 2D marker positions from localization $\langle\!\langle \mathbf{L} \rangle\!\rangle$

Initialize:
Compute/load localization ROI, Ω_{LOC} % *used for initial reduction of image size*

% *for each image in the sequence:*
for $i \leftarrow 1$ **to** $m-1$ **do**

$\mathbf{T}_1$ **Compute tracking prediction:** $\hat{a} = (\hat{t}_u, \hat{t}_v, \hat{\gamma})^T$, Sect. 7.4.3

$\mathbf{T}_2$ **Compute tracking ROI:** Ω_{TR}, using $\hat{a}$ and $b^{\max}$ in Alg. D.5

$\mathbf{T}_3$ **Open and filter fluoroscopic images:** block $\langle\!\langle \mathbf{F} \rangle\!\rangle$, obtain MEF-image $\mathtt{Y}_i$

$\mathbf{T}_4$ **Register images:**

Generate template $\mathtt{W}_{\hat{a}}$, Alg. D.17
Compute point $(u, v, \beta)^T$ of maximal NCC between $\mathtt{W}_{\hat{a}}$ and $\mathtt{Y}_i$, Sect. 6.3
Define optimization starting point a^0 from $\hat{a}$ and $(u, v, \beta)^T$, App. D.5.1

$\mathbf{T}_{4\mathrm{B}}$ Call `fmincon` to solve (6.1.1):

$$(a^*)^i = \arg\min_{a^i} \mathcal{D}_{\mathrm{NNCC}}(\mathtt{W}_{a^i}, \mathtt{Y}_i), \quad \text{s.t. } |a^i_j| \le a^{\max}_j,$$

where:

$\mathbf{T}_{4\mathrm{A}}$ $\mathtt{W}_{a^i}$ is computed in Alg. D.17,
$\mathcal{D}_{\mathrm{NNCC}}$ defined in Sect. 5.2 and App. C.4,
and $a^{\max}$ in Sect. 6.1

$\mathbf{T}_5$ **Save results:** $(a^*)^i$, $(c^k_{2\mathrm{D}})^i = \mathcal{T}^{2\mathrm{D}}((a^*)^i, (c^k_{2\mathrm{D}})^0)$

end

$\mathbf{T}_6$ **Compute 2D marker displacements:** $(d^k)^i = (c^k_{2\mathrm{D}})^i - (c^k_{2\mathrm{D}})^0$

Output: $(a^*)^i$ – optimal transformation parameters, $(c^k_{2\mathrm{D}})^i \in \mathbb{R}^2$ – 2D marker positions, $(d^k)^i$ – marker displacements from the initial positions

Index