IMAGE-BASED PASSIVE ACQUISITION OF RANGE DATA

IMAGE-BASED PASSIVE ACQUISITION OF RANGE DATA

By

SHI XU, B.Eng. M.Eng.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

MASTER OF ENGINEERING (1992) McMASTER UNIVERSITY

(Electrical Engineering) Hamilton, Ontario

TITLE: Image-Based Passive Acquisition of Range Data

AUTHOR: Shi Xu, B.Eng. (Changchun Institute of Optics and Fine Mechanics)
M.Eng. (Xian Institute of Optics and Precision Mechanics, Academia Sinica)

SUPERVISOR: Dr. David W. Capson

NUMBER OF PAGES: xv, 88

ABSTRACT

An image-based technique for passive acquisition of three-dimensional (3-D) range data is proposed. The distance is extracted, in this technique, from the estimation of focus conditions on images produced through a monocular imaging system under natural illumination.

The image taken from a 3-D object is generally out-of-focus (defocused). For each surface point, the severity of defocus on the image depends upon how far away the point is from the imaging system and how camera (optical) parameters are adjusted. Each setting of the parameters can be recorded physically, and associated in object-space with the inverse of a distance that corresponds to the position for the sharpest imaging under this setting. Therefore, for a given surface point the defocus severity is a function of such an inverse object-distance. It can be shown that this function is symmetrical to, and monotonic on both sides of, a point corresponding to the inverse distance of the surface point. To estimate the parameters of the function (one of which is the inverse distance of the surface point), 3~4 images need to be taken under different camera settings with known associated inverse distances in object-space, determined through a once-for-all calibration procedure. Defocus severity is evaluated from a calculation on the window image that corresponds to a

small area around the surface point, and the inverse variance in the window is suggested in this technique for the best performance. The 3-D surface geometry is acquired by applying the algorithm, in parallel, to all surface points in the field of view.

Various aspects of the technique are discussed and several algorithms are developed. The technique is implemented on an opto-digital imaging system and evaluated under different conditions. A number of objects are tested to demonstrate its performance.

ACKNOWLEDGEMENTS

The patience and guidance from my supervisor Dr. David Capson is mostly appreciated. The financial assistance from Natural Sciences and Engineering Research Council of Canada and McMaster University is acknowledged. I am also thankful to my friends at McMaster, particularly those in Image Analysis Laboratory, with whom I have enjoyed my stay at McMaster.

I dedicate this thesis to my family. Their constant support and encouragement have all made it possible.

TABLE OF CONTENTS

LIST OF FIGURES & TABLES

LIST OF SYMBOLS & ABBREVIATIONS

δ	Error of Measurement.
$\Delta l'$	Depth of Focus.
Δl	Depth of Field.
θ,ϕ,r	Orthogonal-Axis Coordinates.
λ	Optical Wavelength.
σ	Standard Deviation.
ω	Angle of Field of View.
2-D	Two-Dimensional.
3-D	Three-Dimensional.
A	Window Area.
a	Constant.
B	Area in Spatial Frequency Plane.
b	Constant.
C_i,c	Constant.
CCD	Charge-Coupled Device.
d	Image Blur Diameter.

D Aperture Width (Diameter).

e Difference between Distances.

$\mathscr{E}(\ )$ Energy.

Err. Distance Deviation.

F F-number, f/D.

f Focal Length.

$\mathscr{F}(\ )$ Fourier Transform.

f( ) Intensity Distribution.

f_0 Incoherent Cutoff Frequency.

f_m Average Intensity.

$(f^2)_m$ Mean Square of Intensity.

$f_x f_y$ Spatial Frequency Components.

FFT Fast Fourier Transform.

FSC Focus Sharpness Criterion (Criteria).

h Object Height.

$\boldsymbol{h}$ Height Vector.

I_i Image.

$\boldsymbol{i,j}$ Orthogonal Unit Vectors.

J( ) Focal Gradient Function.

J_0 Parameter in Hyperbola Algorithm.

J_i Value of J( ).

J_{ij} Increment of J( ).

K	Scale Factor.
k	Constant of Proportionality.
L	Sampling Interval.
l,l_i	Object-Distance.
l'	Image-Distance.
l_0	Distance of Surface Point.
l_0'	Distance of Surface Point Image.
M	Image Size in Pixel.
m,m_i	Magnification.
n	Number of Images.
n'	Refractive Index in Image-Space.
N	Image Size in Pixel.
O,O^*	Optical-Axis Location.
O_i	Axial Point in Object-Space.
O_i'	Axial Point in Image-Space.
OTF	Optical Transfer Function.
P( )	Power Spectral Density.
p_i	Records of Camera Parameters.
P_i	Object-Plane.
P_i'	Image-Plane.
PSF	Point Spread Function.
Q_0	Object Point.

Q_0' Image Point.

Q_i Projection of Q_0.

Q_i' Image of Q_i.

R,R_i Normalization Factor for Image Scale.

S,S_i Feature Point.

SfD Shape-from-Defocus.

SfF Shape-from-Focus.

T Normalization Factor for Inverse Distance.

t,t_i Inverse Distance.

t Normalized Inverse Distance.

t_0 Inverse Distance of Surface Point.

U' Half Aperture Angle in Image-Space.

V( ) Grey-Level Variance.

W,W_i Window Size in Pixel.

x Independent Variable.

X,Y,Z Cartesian Coordinates.

x,y,z Values of X,Y,Z.

CHAPTER 1

INTRODUCTION

1.0 Introduction

A primary goal of a visual system is to perceive and acquire information about the 3-D world. In machine vision, the knowledge of a scene, especially its 3-D geometry, is often crucial for supporting tasks such as scene recognition, automated inspection, and mobile robot navigation (Horn, 1986; Kanade, 1987; Shirai, 1987).

Range data acquisition is always an interesting, and sometimes challenging field to explore. During recent years, a great deal of effort has been made in the computer vision community, and in industry at large, to extract this important information about the world around us (Besl, 1988). In this chapter, we present some basic concepts concerning the acquisition of 3-D range information and review briefly some related work in the literature.

1.1 Image-Based Ranging and Range Image

A variety of ranging techniques that collect 3-D coordinate data of visible object surfaces in a scene are now available varying from noncontact optical methods

to those based on tactile sensing (Jarvis, 1983; Shirai, 1987; Besl, 1988). Amongst these noncontact techniques, *image-based* approaches, which acquire 3-D (depth) information from 2-D (intensity) image(s) of a scene, are particularly useful where *geometric* (scene depth) data need to be extracted in parallel (without physical scanning), or *photometric* (scene radiance) information is required at the same time (Jarvis, 1983). Shape-from-shading, depth-from-texture, and depth-from-motion, for example, can be categorized as image-based.

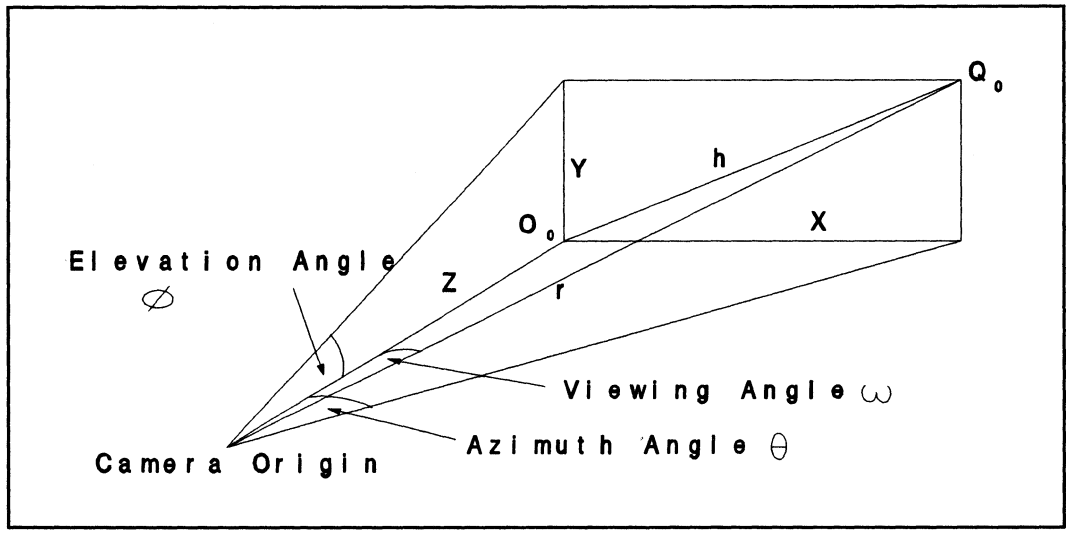


Figure 1.1 Coordinates with Origin at Camera.

The acquired geometric data are described as *range image*, also known as a range map or depth map. The term "image" is used here because a range image can be displayed on a video monitor, in the form of a digitized video image. Generally speaking, a range (geometric) image and the corresponding photometric image are in common in that a scene point can be "imaged" at the same location on both images. Apparently, they are different in that the value of a pixel on the range image gives depth information, while that on the photometric image represents

brightness. Depending on how the distance is measured, the resultant range image is usually in the form of Cartesian *Z-XY*, or Orthogonal-Axis *r*-θφ, or the mixed *Z*-θφ coordinates (Besl, 1988). In the coordinates system with origin at the camera, as shown in Figure 1.1, these coordinates are related through:

$$\begin{aligned} x &= \tan\theta \cdot z \\ y &= \tan\phi \cdot z \\ r &= \sqrt{1+\tan^2\theta+\tan^2\phi}\ \cdot z \end{aligned} \qquad (1\text{-}1)$$

Most optical imaging systems are axially symmetric. Unless otherwise specified, the optical axis is always chosen as *Z*-axis. Under such a coordinate system, referring to Figure 1.1, the coordinates of an object point Q_0 are related to its height h and viewing angle ω by:

$$\boldsymbol{h} = x\boldsymbol{i} + y\boldsymbol{j} \quad or \quad \tan^2\omega = \tan^2\theta + \tan^2\phi \qquad (1\text{-}2)$$

where $\boldsymbol{h}$ is the height vector, and h represents its magnitude.

Figure 1.2(a) is an example of range image from our experiment, the depth map of two spring washers; the corresponding image of brightness and 3-D surface plot are shown in (b) and (c) respectively.

1.2 Focus-Based Techniques

Most of the image-based techniques are based on a pin-hole camera model, *i.e.*, the image from a camera of finite aperture size, which is generally defocused or blurred (except for some in-focus points), will be considered as "imperfect".

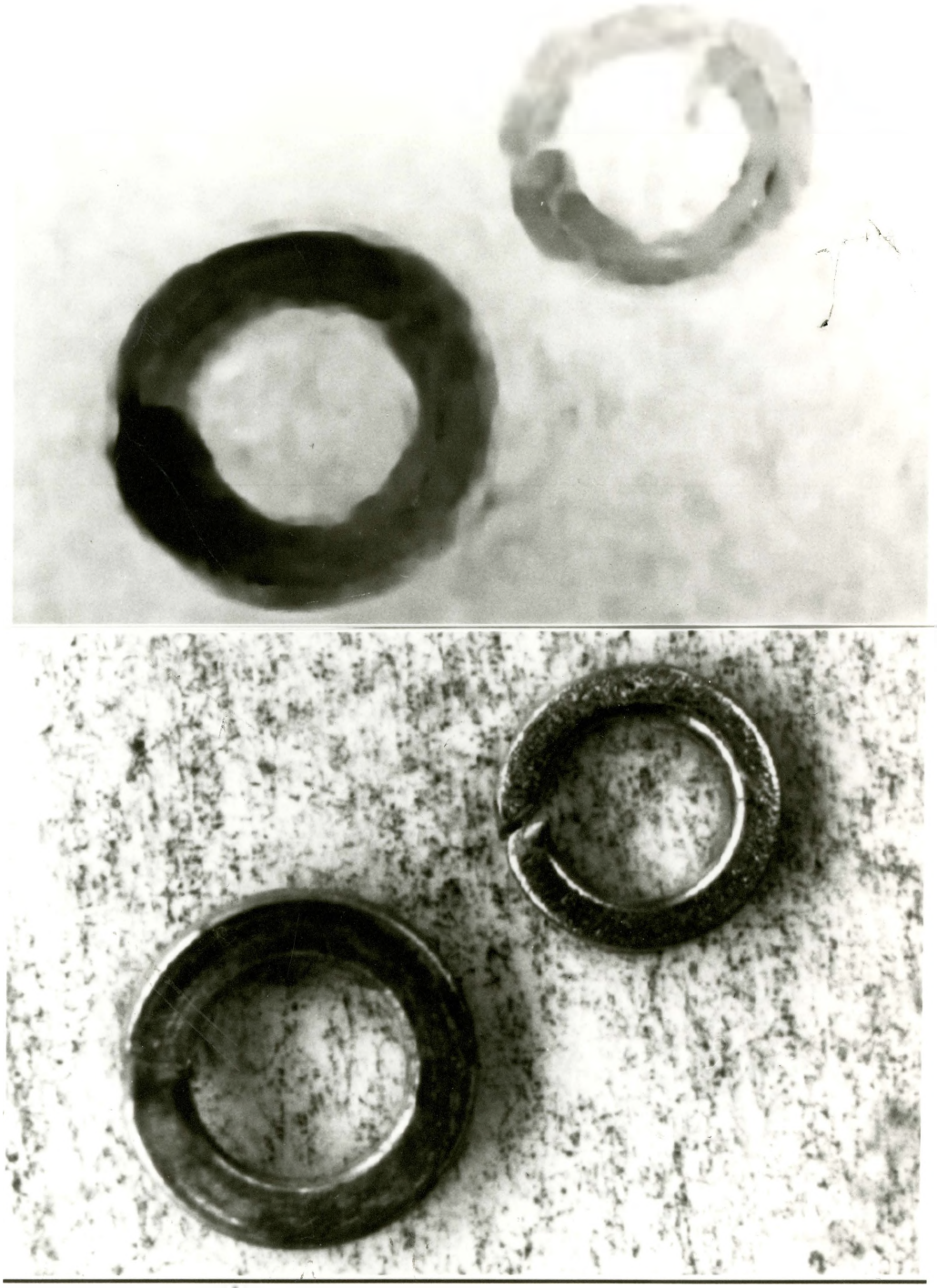


Figure 1.2 Example: Spring-Washers.
(a) Depth Map.
(b) Image of Brightness.

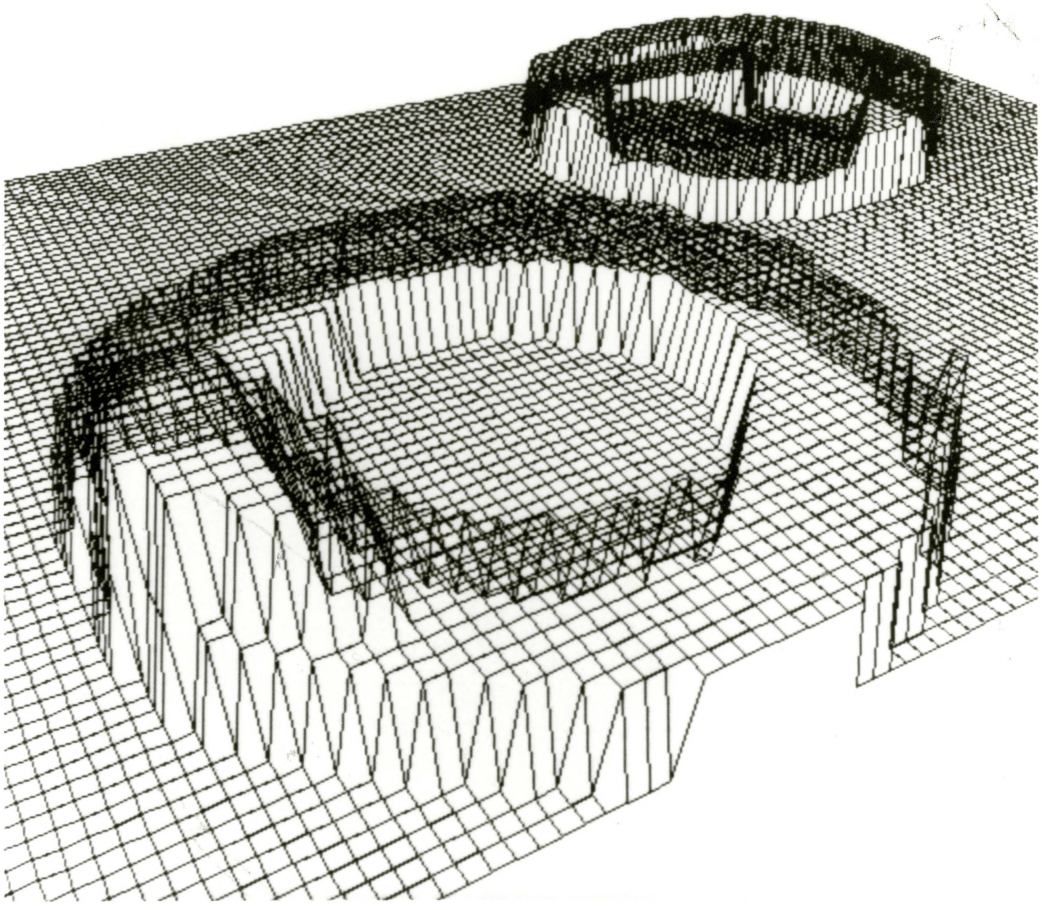


Figure 1.2 Example: Spring-Washers.
(c) 3-D Surface Plot (skeleton).

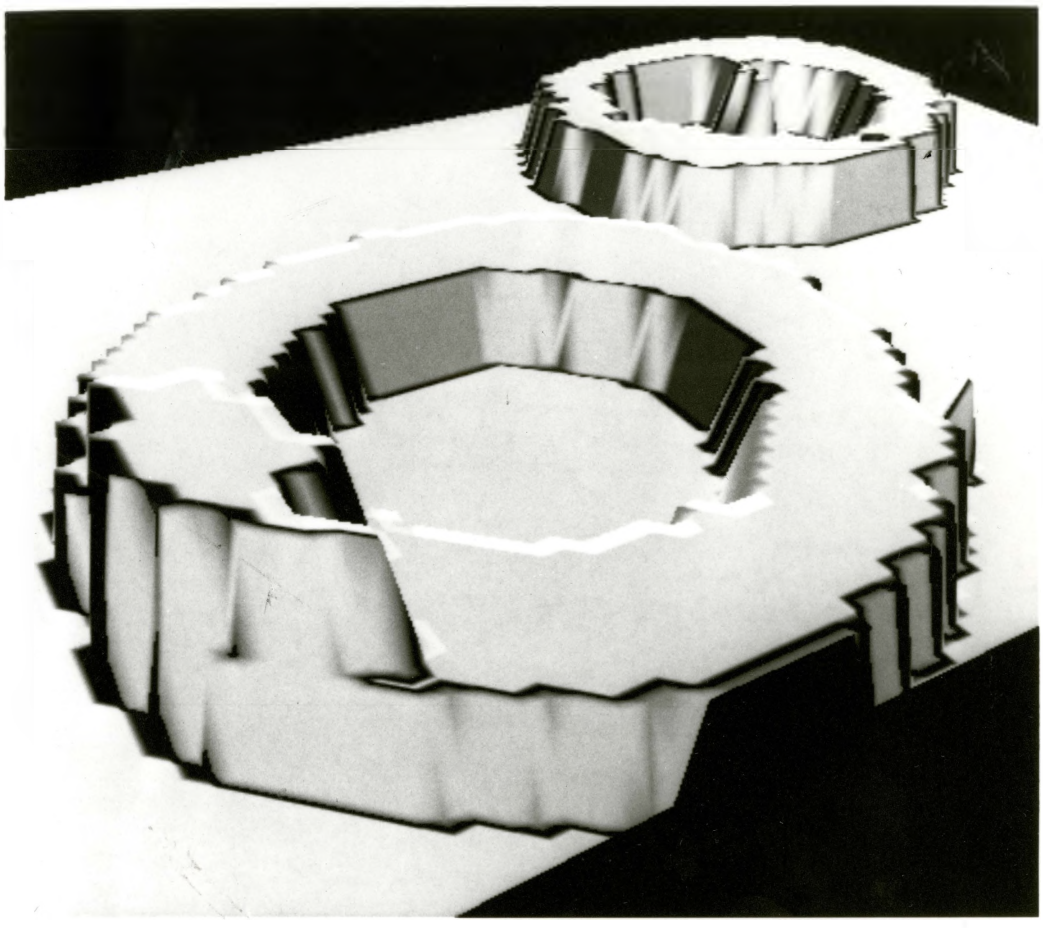


Figure 1.2 Example: Spring-Washers.
(c) 3-D Surface Plot (rendered).

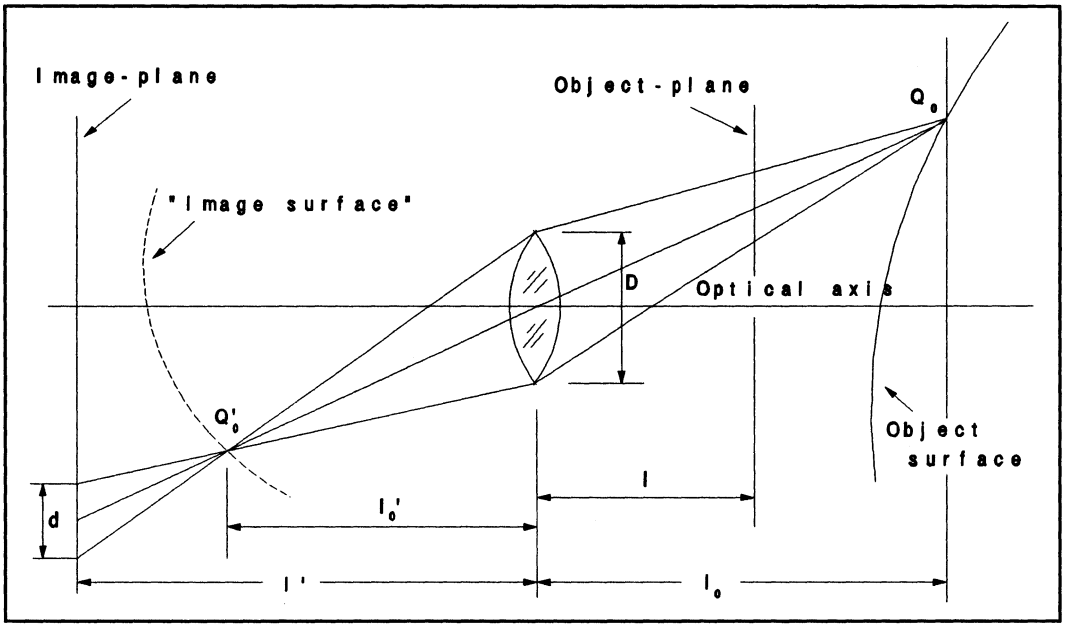


Figure 1.3 Image Formation and Defocus.

However, the "imperfect" image is itself a useful source of depth information. As shown in Figure 1.3, the surface of a 3-D object is "copied", through the imaging system, into an "image surface" on which all the surface points are exactly focused. The image taken at l' is generally defocused. The geometrical blur on the image-plane, d, is related to the distance l_0 of the surface point Q_0 by:

$$l_0 = \frac{f \cdot l'}{l' - f \pm dF} \qquad (1\text{-}3)$$

where f is the focal length, and F refers to the system's F-number, f/D. Note that l' and l_0 in Eq.(1-3) are related, respectively, to distances l and l_0' in Figure 1.3 by Lens Law:

$$\frac{1}{l} + \frac{1}{l'} = \frac{1}{f} \qquad \textit{and} \qquad \frac{1}{l_0} + \frac{1}{l_0'} = \frac{1}{f} \qquad (1\text{-}4)$$

Image-based methods that exploit this kind of information are *focus-based*, and are often referred to as shape-from-(de)focus. This approach avoids, inherently, heuristic assumptions on the scenes and images (Subbarao, 1989; Hwang, 1989), which are often used in techniques based on the pin-hole camera model. For instance, in depth-from-stereopsis the correspondence problem often requires heuristic solutions; in shape-from-shading, the reflectance model of visible surfaces needs to be assumed in order to recover the surface geometry. The problem of image point-to-point correspondence does not exist in this approach, and the regional correspondence between images can be attained, in general, by a simple magnification modification or normalization (Horn, 1968; Subbarao, 1989).

1.2.1 Shape-from-Focus

To estimate the quality of focus, a number of *focus sharpness criteria* (*criterion*), abbreviated as FSC, have been proposed (Jarvis, 1983; Krotkov, 1987; Nayar, 1992). Since σ (refer to Figure 1.3) exhibits its global minimum (ideally zero) if $l' = l_0'$, where l_0' is the image-distance corresponding to l_0, the FSC must be so designed that the global extreme value is assumed at l_0. Obviously, searching for the FSC extreme by constantly accommodating camera parameter(s), usually the image-plane position, is a way to find the object-distance of a surface point. Techniques employing this strategy are referred to as *shape-from-focus* (SfF).

Although fast searching algorithms can be applied (Krotkov, 1987), this approach is inherently slow since it involves recording and computing a large number

of (theoretically infinitive) images. Horn (1968), Jarvis (1976, 1983) provided some technical details on focusing relationship and computational formulae for FSC. Krotkov (1987) evaluated and compared different FSC. Das (1989) presented an interesting approach integrating both stereo and focus as sources of depth for surface reconstruction. Schlag (1983), Engelhardt (1988), Darrell (1988), and Nayar (1992) also presented different approaches based on SfF strategies.

1.2.2 Shape-from-Defocus

Instead of accommodating the vision system for the best focus, Pentland (1987, 1989), Rioux (1986), and Grossman (1987) proposed and implemented algorithms based on so called *shape-from-defocus* (SfD) (Hwang, 1989).

In their algorithms, only a few (even one) images are processed and the distances are estimated from defocus conditions on the images. It has also been investigated that the human visual system may use the same information, the gradient of defocus, to perceive depth (Pentland, 1987). Subbarao (1987, 1988, 1989) presented more general solutions for SfD. Hwang (1989), and Lai (1992) further modified and generalized Pentland's algorithms with some experimental results. Cardillo (1991) presented a calibration scheme for such SfD methods.

Most of the focus-based techniques are *passive*, by which we mean that no controlled illumination is involved, and thus the natural scene radiance is recovered along with the depth map. Two exceptions are the techniques given by Engelhardt (1988) and Rioux (1986), where controlled grating and grid illuminations, and

aperture masks were used.

1.3 Thesis Organization

In this thesis, we are concerned with passive SfD techniques. In Chapter 2, such techniques are reviewed and a new technique is proposed and discussed. This new technique acquires depth information by modelling the gradient of focus and both scene depth and albedo can be recovered in parallel. The technique is detailed in Chapter 3, which includes the related rationale and algorithms. An implementation of a system is described in Chapter 4 together with some experimental results. A discussion of the technique and recommendations for further work conclude the thesis in Chapter 5.

CHAPTER 2

PASSIVE SHAPE-FROM-DEFOCUS

2.0 Introduction

In this chapter, we first discuss, in general, passive shape-from-defocus (SfD) techniques. A new technique based on modelling the gradient of defocus in image-space is then introduced. Emphasis is placed on the comparisons amongst the assumed models in these techniques.

2.1 Modelling Point Spread Function

The blur size d in Eq.(1-3) is directly related to the *point spread function* (PSF) with respect to the image position in the imaging system, and it is often referred to as the *spatial constant*, or *spread parameter*, of the PSF. Apparently, if we can somehow estimate d, the object-distance l_0 can be derived from Eq.(1-3). Indeed, most passive SfD techniques are based on modelling or evaluating the PSF, and Eq.(1-3) or similar formulae are more or less involved. These techniques appear to be practically feasible, and results with reasonable accuracy have been reported (Grossman, 1987; Pentland, *et al.* 1989; Cardillo, 1991; Lai, 1992). However, such

PSF model-based approaches are inherently inaccurate, and thus do not ensure high quality results (Nayar, 1992).

Since Eq.(1-3) is deduced from the geometry of lens imaging, d in the equation only represents the pure geometrical blur caused by out-of-focus, or defocus. On the other hand, the actual blur, whatever means is used to obtain it, is not caused by geometrical defocus alone. Therefore, substituting d in Eq.(1-3) with the actual blur would yield errors (usually nonlinear) in calculating distance l_0. Replacing d by kd (Pentland, 1987; Subbarao, 1988), where k is a constant of proportionality determined through camera calibration, may eliminate the effect of diffraction, but not the effect from factors such as optical aberrations, vignetting, discretization, *etc.* In the worst case, where an object point happens to be exactly in-focus, d is zero but the actual blur, caused by aberrations and other effects, is not.

Secondly, in these techniques, d must be characterized as one of the parameters in a PSF model. For mathematical simplicity, a bivariate symmetric Gaussian was suggested (Pentland, 1987), and d was taken as its spatial constant. According to Central Limit Theorem, the model is accurate where defocus is comparable to many other blur factors, which is often the case where defocus is very small. Obviously, the model is inaccurate where defocus prevails. In this case, the geometrical-optics predicted model, the symmetric cylindric one, would be more appropriate for describing the PSF (Goodman, 1968). Subbarao and Natarajan (1988) even provided a more general alternative: the circularly symmetric model. The second order central moment is taken as the spatial constant. However, in a

practical optical imaging system, which is not shift-invariant, the symmetrical PSF is not an exact model for off-axis image points. Also, in the early stage of processing, the differentiation operation (for central moments) is often vulnerable to noises. In an attempt to generalize an algorithm using defocused edges to infer depth, Lai *et al.* (1992) decomposed the spatial constant in Gaussian PSF into two orthogonal ones. Though less sensitive to noise and edge orientation, the algorithm is based on the assumed validity of the Gaussian model.

Indeed, because of the complexity involved, it is difficult to model PSF that is both analytically accurate and technically feasible. In the next section, a new method for determining SfD without modelling PSF is described.

2.2 Modelling Gradient of Focus

An alternative is to model the gradient of focus (or defocus). As shown in Figure 1.3, the size of the blurred image of a surface point Q_0, d, will vary as we make certain changes on camera (optical) parameters, for example, by moving the image-plane back and forth (changing the parameter l'). In other words, a gradient of (de)focus exists in image-space. Since such a change on camera parameters also gives rise to the change of object-plane position, the position from which an object is sharply focused, we can always associate this gradient of de(focus) in image-space with such a positional change in object-space. For the simple lens system in Figure 1.3, for example, we can easily associate d in image-space with the object-distance l from Eq.(1-2) and (1-3):

$$d = \left|\frac{t_0}{t} - 1\right| \frac{f}{F} \cdot m \qquad (2\text{-}1)$$

where $t = l^{-1}$ is the inverse object-distance; $t_0 = l_0^{-1}$ refers to the inverse distance of the surface point Q_0; and

$$m = \frac{l'}{l} \qquad (2\text{-}2)$$

is the magnification associated with each pair of object/image planes.

2.2.1 Focal Gradient

Conceptually, the gradient of focus and the gradient of defocus are different: "focus" here implies the quality of focus, and should be so quantitatively described that its global *maximum* is taken at the position of exact focus (Jarvis, 1983); "defocus", on the other hand, refers to the severity of defocus, and is evaluated so that a global *minimum* occurs at the same position. The blur size d on the image plane, for example, can be a measure of defocus in this sense since it takes its global minimum, which is zero ideally, at the position of exact focus. Obviously, a measure of focus quality, if used inversely, can be a measure of defocus severity. In this light, the *focal gradient* is used, in this thesis, to refer to either the gradient of focus or the gradient of defocus, and, as we already do in Chapter 1 without explanation, the focus sharpness criterion (FSC) is used in its broader sense (Krotkov, 1987), *i.e.*, a FSC assumes its global *extreme*, either maximum or minimum, at the position of exact focus.

2.2.2 Parameter Estimation from Focal Gradient

In practice, the focal gradient is quantitatively evaluated with a FSC. The *focal gradient function*, which describes the aforementioned relationship between the focal gradient and the object-plane position, is then defined as a function of normalized FSC in terms of the inverse object-distance t, and will be represented by $J(t)$ throughout the thesis. Without question, $J(t)$ must also assume a global extreme at $t=t_0$, where lies the surface point Q_0. In addition, it will be shown that $J(t)$ is symmetrical to, and monotonic on both sides of, the point t_0.

Apparently, the geometrical centre of the $J(t)$ curve indicates the position of exact focus. Mathematically, this centre, namely t_0, can be estimated from only finite number of $J(t)$ values at known points provided an analytical model for the curve is assumed. In the proposed technique, the acquisition of range data is, in essence, such an estimation procedure for the parameter t_0 from $J(t, t_0)$.

Applying this procedure to every point on the images, we are able to extract the distance for every surface point and thus establish the whole surface geometry. With parallel processing technology and the programmable motion control system currently available, the distances can be extracted in parallel, and in real-time.

2.2.3 Parameter Association through Calibration

In this parameter estimation technique, the value of $J(t,t_0)$ at a point, say $J(t_i,t_0)$, $i=1, 2, \ldots \ldots, n$, is acquired from the ith image taken with a known setting of camera parameters. The number n here refers to the number of points required to

estimate t_0, which is equivalent to the number of images needed. A calibration procedure, which determines the relationship between the camera setting and t_i, is required for the following reasons:

1) In general, it is difficult to obtain an accurate analytical relationship between a setting of camera parameters and t in object-space;

2) Even if such a relationship does exist, as is the case for a simple lens system, where one of the camera parameters, l', is analytically associated with $t=l'^{-1}$ through Eq.(1-4), it is always less accurate to infer t_i from such a relationship than physically searching for the best association.

3) In practice, camera parameters are often better recorded and acquired physically in terms of their changes, *i.e.*, their relative values. In this case, analytical relations such as Eq.(1-4) cannot be applied.

In addition, through the calibration, the magnifications between each associated pair can also be found so that the magnification normalization can be carried out accurately for the regional correspondence amongst the images. Though it is time-consuming and computationally complex, the calibration is a once-for-all procedure.

CHAPTER 3

DEPTH ESTIMATION FROM FOCAL GRADIENT

3.0 Introduction

In this chapter, we elaborate this new technique that estimates depth from the focal gradient. We first address the problem of regional image correspondence; then discuss, in detail, the validity of the assumed attribute for the focal gradient function and how a FSC, the *energy measure* in particular, can be applied for describing the function. Finally, some algorithms based on different analytical approximations for the focal gradient model are presented.

3.1 Regional Correspondence

The diagram of a monocular camera system is given in Figure 3.1. The object-planes at different positions, P_1, P_2, ..., P_n, are associated, respectively, with image-planes P_1'~P_n'. Depending upon the system and how the camera parameters are adjusted, the image-planes may physically coincide. For the object-planes, the differences in position result from the changes of camera parameters. The physical movement of the image-plane, the translation of the lens along the optical axis, and

the change of focal length (for the zoom lens), for example, are all considered in this technique as such parametric changes.

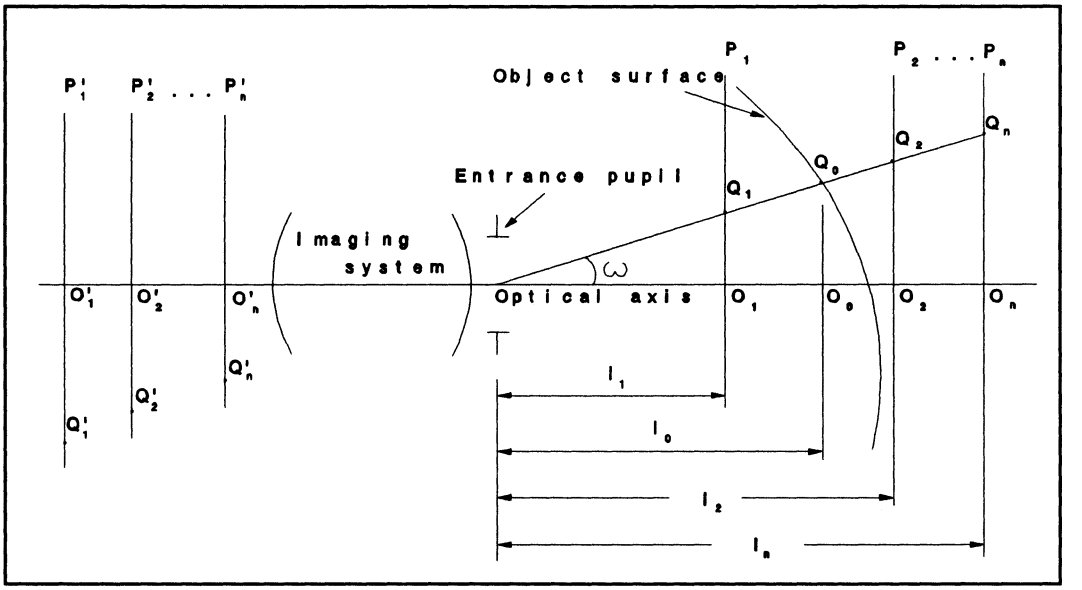


Figure 3.1 Geometry of Image Correspondence.

3.1.1 Scale Normalization

The points Q_1, Q_2, ..., Q_n on the object-planes are the geometrical centres of the projections from Q_0, the object surface point. Q_1', Q_2', ..., Q_n' in image-space are the correspondents to those projection centres in object-space. By the *correspondent* we mean that if an object point is positioned at Q_i (i=1, 2, ..., n), an exactly focused image of the point must occur at Q_i' in image-space. As mentioned before, the association of each acquired image P_i' with the inverse distance t_i and the magnification between the associated object/image pair, m_i, are determined through a system calibration procedure.

Although the acquired images usually have the same size, and there is no

rotation among them, they are often different in scale. In Figure 3.1, for example, we usually have $Q_1'O_1' \neq Q_2'O_2' \neq \ldots\ldots \neq Q_n'O_n'$. In other words, there is a region-to-region matching problem. Although the problem was noted in some of the previous work (Krotkov, 1987; Subbarao, 1989; Nayar, 1992), there appears to have been no experimental result where it was both addressed and solved.

First of all, the images must be normalized so that they are all the same in scale. In this technique, given a calibrated system (with known t_i and m_i associated with each image), the normalization procedure is straightforward. From the geometry in Figure 3.1, we have

$$Q_i' O_i' = m_i \cdot l_i \cdot \tan\omega \qquad (i = 1, 2, \ldots, n) \qquad (3\text{-}1)$$

where $m_i = Q_i'O_i'/Q_iO_i$, determined through the calibration; $\tan\omega = Q_0O_0/l_0$.

Dividing $Q_i'O_i'$ in Eq.(3-1) by a factor

$$R_i = \frac{m_i \cdot l_i}{K} \qquad (3\text{-}2)$$

yields the normalized image height that only depends on ω, the viewing angle corresponding to Q_0:

$$h = K \cdot \tan\omega \qquad (3\text{-}3)$$

where K is a scale factor which determines the actual sizes of the normalized images.

In practice, K is often chosen so that the values of R_i are close to 1. For a given digital image resolution, selecting $R_i < 1$ results in images becoming larger and we gain no extra information. For $R_i > 1$, images become smaller and we lose

information.

As shown in Eq.(1-2), given the directions of two orthogonal coordinates, the height vector $\boldsymbol{h}$ whose (normalized) magnitude is described with Eq.(3-3) can always be decomposed into two independent components. Therefore, depth maps obtained from this technique must be in the form of z-$\theta\phi$ or z-tanθtanϕ, which, according to Eq.(1-1), can be directly converted into the more common Cartesian system.

3.1.2 Reference Point

Note that to perform only scale normalization, it is not necessary to assume a reference point or the origin of coordinates on the images. However, it is necessary to locate at least one reference point on the images that have been normalized, to ensure a complete matching amongst them. Feature points on the images are used to establish the correspondence among images in shape-from-stereopsis. In SfD, a feature point could also be taken as the reference point on the images, though finding such a feature point may not be as straightforward as normalizing the scale and a blurred feature point on the defocused images may introduce nonnegligible matching errors. The regional correspondence problem could then be solved and a depth map be acquired. However, without knowing the location of the optical axis, it is impossible to convert the depth map into real world 3-D coordinates.

The position of the optical axis on the images can be determined through a once-for-all calibration procedure and, within a distance range, it can be accurately

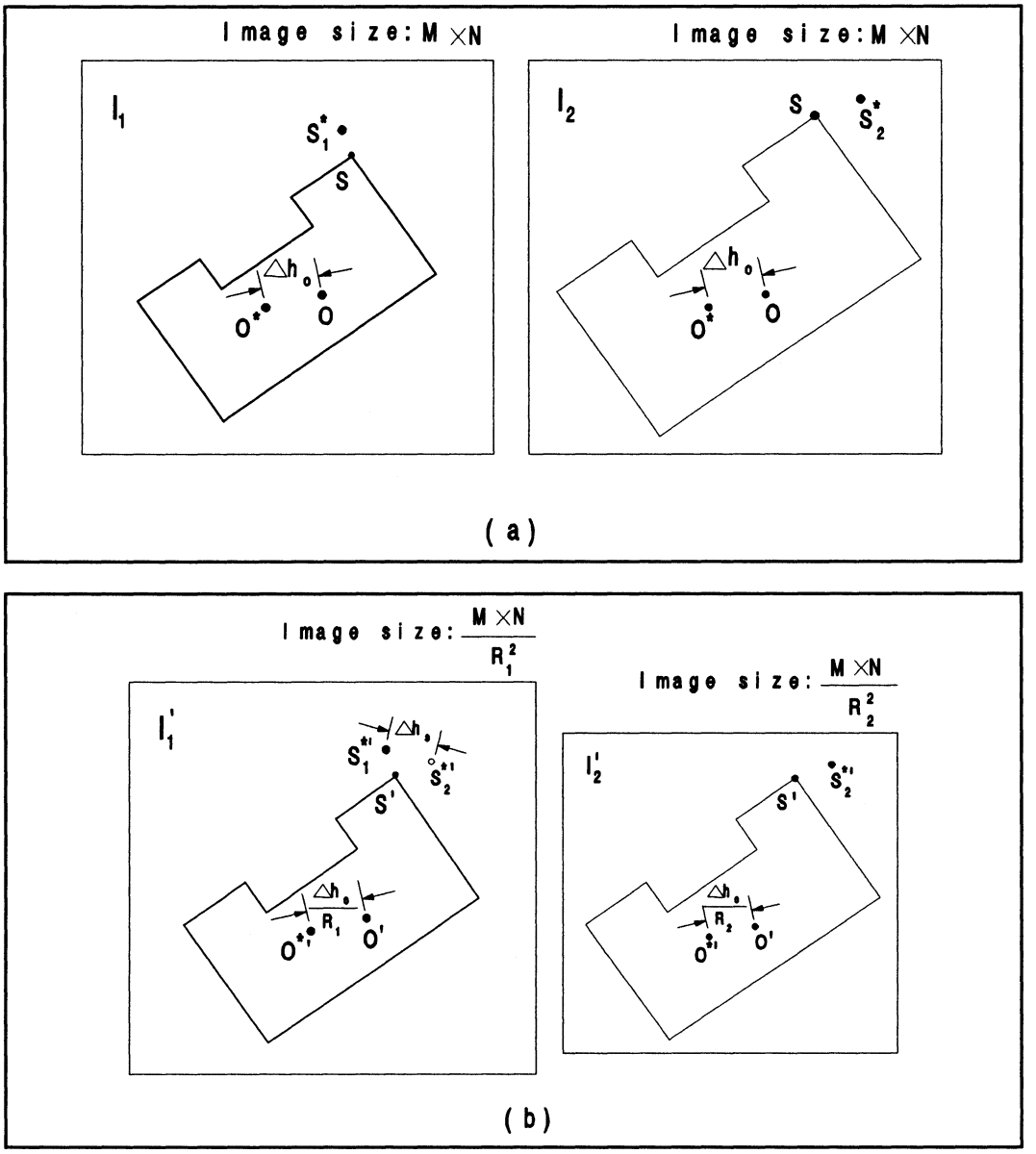


Figure 3.2 Feature Points and Optical-Axis.

located using readily available optical methods. More importantly, the optical axis on the images is itself a good reference point for image correspondence. It can be shown that the matching error caused by the error of locating the optical axis on the images is far less than that caused by the error of locating a feature point.

Suppose that two M×N images in Figure 3.2(a), I_1 and I_2, are taken for the

same object with different l_i, and thus have different scales. The normalized images, I_1' and I_2', are in Figure 3.2(b), where they become the same in scale but different in size. On the original images, O represents the real position of the optical axis which is, say, inaccurately located (in a once-for-all calibration) at O^* and the error is $|\Delta \boldsymbol{h}_o|$, where $\boldsymbol{h}_o$ is the height vector for the optical axis position on the images. Suppose S then represents a feature point, which is erroneously located (in a measurement) at S_1^* and S_2^* on I_1 and I_2 respectively. O', O^*', S', S_1^*' and S_2^*' are the corresponding points on the normalized images, and $|\Delta \boldsymbol{h}_s|$ represents the relative error locating S', where $\boldsymbol{h}_s$ is the height vector for the feature point. R_1 and R_2 in the figure are, referring to Eq.(3-2), the normalization factors for I_1 and I_2 respectively.

Obviously, if S were chosen as the reference point, two normalized images would be dislocated to each other by $|\Delta \boldsymbol{h}_s|$. On the other hand, if we choose the location of optical axis on the images as the reference point, the images will only be dislocated by $|\Delta \boldsymbol{h}_o| \cdot |1/R_1 - 1/R_2|$. In SfD, the differences of images in scale are usually very small, and so is $|1/R_1 - 1/R_2|$. In our experiment, for example, five object-planes are calibrated, and $|1/R_1 - 1/R_5|$ is only about 0.015 (assuming $R_3 = 1$).

Applying this normalization procedure in both orthogonal directions to all points on the images will ensure the direct matching amongst all images, region to region, point to point. Note that the position of the entrance pupil is implicitly assumed to be fixed in deriving Eq.(3-3). Indeed, this condition can be readily ensured in practice, and it enables all the projections from Q_0 to be in the same

direction toward the entrance origin so that a linear scale modification is made possible. Subbarao (1988) noticed that the lens should not be moved during the measurement. Further analysis shows that in a general system, except for images of objects of no depth variation (the vertical planes), fixing the entrance pupil position is the only way to ensure the correspondence amongst the images.

3.2 Focal Gradient Function

The symmetry and monotonicity is the necessary attribute assumed for J(t), the focal gradient function. A real J(t) curve obtained from this experiment is shown

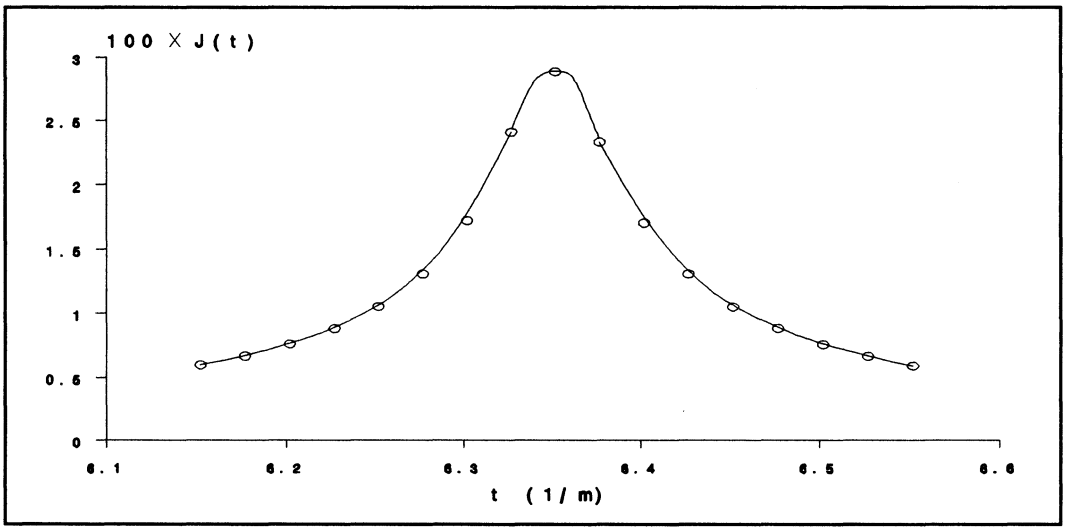


Figure 3.3 Focal Gradient Function.

in Figure 3.3 (through aperture of F/2.8). In this model, the focal gradient is practically evaluated by a focus sharpness criterion (FSC). We first examine how the FSC in a *diffraction-limited* system behaves with respect to t. Obviously, the only cause for image degradation in such a system is out-of-focus. In other words, for

such a system, the *quality of image* and the *quality of focus* imply the same, as do a criterion for image quality and a FSC. Since the quality of image in a system is primarily characterized by its point spread function (PSF), any criterion for image quality, a FSC in this case, should behave, in principle, in agreement with the PSF.

A simple geometrical relationship in Eq.(2-1) shows how the PSF, characterized here by its spatial constant d, varies with t. After the necessary magnification normalization on the images by R in Eq.(3-2), we have

$$d = \frac{f}{F} \cdot |t - t_0| \qquad (3-4)$$

which is symmetrical to and monotonic on both sides of t_0, and has its minimum at this position.

The analysis based on diffraction optics provides in the spatial frequency domain a similar conclusion. The *optical transfer function* (OTF) for the diffraction-limited system of a square aperture is derived as (Goodman, 1968, p.125):

$$\mathrm{OTF}(t-t_0, f_x, f_y) = \mathrm{tri}\left(\frac{f_x}{f_0}\right) \cdot \mathrm{tri}\left(\frac{f_y}{f_0}\right) \cdot \mathrm{sinc}\left[\frac{(t-t_0)D^2}{\lambda}\left(\frac{f_x}{f_0}\right)\left(1-\frac{|f_x|}{f_0}\right)\right] \cdot \mathrm{sinc}\left[\frac{(t-t_0)D^2}{\lambda}\left(\frac{f_y}{f_0}\right)\left(1-\frac{|f_y|}{f_0}\right)\right] \qquad (3-5)$$

where λ is the wave length; f_x and f_y are two independent spatial frequency components; D is the aperture width.

The incoherent cutoff frequency, beyond which no frequency component is transmitted through the system, is defined as:

$$f_0 = \frac{D}{\lambda \cdot m \cdot l} \qquad (3-6)$$

where m, referring to Eq.(2-2), is the magnification associated with the imaging position.

After the images are normalized by R in Eq.(3-2), f_0 becomes a constant, *i.e.*, it is irrelevant to $t = 1/l$. The functions tri(x) and sinc(x) in Eq.(3-5) are defined as:

$$\begin{aligned} \mathrm{tri}(x) &= \begin{cases} 1-|x| & |x| \le 1 \\ 0 & otherwise \end{cases} \\ \mathrm{sinc}(x) &= \frac{\sin(\pi x)}{\pi x} \end{aligned} \qquad (3-7)$$

At a given frequency point (f_{x0}, f_{y0}), OTF is only a function of t-t_0. This function is symmetrical to and maximizes at t_0. Due to the minor fluctuations of sinc(x) beyond its first zero, OTF(t-t_0) is not monotonic in general. However, if the defocus is kept within a range (so that x in sinc(x) is not beyond the first zero), the monotonicity can be assured. This is the case for this technique since only certain (usually small) depth can be extracted with the technique. Moreover, in most applications, not just one but a range of spatial frequency components are involved. The integration of OTF over such a range is also a function of t-t_0 and this function, on close examination, is monotonic on both sides in general.

It is worth noting that OTF in Eq.(3-5) is the normalized form of the Fourier transform of PSF. Since the magnitude of brightness on images varies as we change the camera parameters, the non-normalized transfer function, or any other non-normalized criterion for image quality, may not have the desired properties. This

intensity normalization for image quality criterion is not critical in shape-from-focus (SfF) (Krotkov, 1987), since the position of the global extreme of a focus sharpness criterion (FSC) is little affected by the brightness variation. In fact, most FSC in the literature are defined in non-normalized forms. However, the symmetry of J(t) will be affected by this variation of intensity with camera parameters, and, in this sense, those criteria are not properly presented.

For the sake of mathematical simplicity, a square aperture is assumed. An examination of PSF for the system with a circular aperture (Born, 1965) would give the same result. Indeed, the conclusion from this analysis applies in general to any aperture shape.

The real optical imaging system is not diffraction-limited. The presence of sources other than defocus in the system, which contribute to the degradation of image quality, will inevitably affect the shown properties of J(t). The aperture of the system often plays a key role in controlling the non-focus sources on the image degradation. For example, according to the theory of primary aberration, given the view angle ω, the spherical lateral aberration is only proportional to D^3, the aperture to the third, and the coma aberration, to D^2. Generally speaking, if the aperture is kept unchanged during a measurement, the "non-focus" effect is practically constant over a certain range of t. Obviously, this constancy may not hold well for t over a large range. However, over such a range, defocus must prevail, and the comparatively small "non-focus" effect is often negligible. Therefore, the symmetry and monotonicity of J(t) should remain in a practical system, as is evidenced from

the real J(t) curve in Figure 3.3 where the normalized *energy* is used to evaluate the function.

3.3 Focus Sharpness Criterion

In principle, J(t) can be evaluated with any FSC. Krotkov (1987) listed and discussed most FSC that had appeared in the literature. All the criteria are consistent in that the global extremes are assumed at the sharpest focus.

3.3.1 Window Operation

For each point on an image-plane, Q_i' in Figure 3.1 for example, a FSC is usually evaluated with a calculation from a window on the image-plane, centred at Q_i'. Windows centred at Q_1', Q_2', ..., and Q_n' all correspond to the same small area on the object surface, centred at Q_0. This window-to-window correspondence is ensured, in principle, with the ensured regional correspondence (Section 3.1).

Apparently, this technique only acquires the average distance within the small surface area. Eventually, the acquired 3-D surface geometry is a "flattened" version of the real surface — the rapid (usually minor also) depth variation on the surface diminishes in the meantime. The larger the window, the more the surface is flattened.

On the other hand, the accuracy of distance estimation in this technique depends partly on the accuracy of the calculation from the window, which is, in turn, affected by factors such as random noise, window position error on the image,

sampling and grey-level quantization, and even the abundance of spatial frequencies from the intensity distribution in the window. With the window large enough, all of the effects can be statistically reduced or even eliminated. In this sense, the larger the window, the better the accuracy of the acquired distances.

Therefore, a balance between the two requirements is necessary. Ideally, the window should be such in size that the resultant distance error from such a window calculation matches the actual depth variations in the small surface area corresponding to that window. We can only select the proper window size for a specific application by trial-and-error. Generally speaking, larger windows for smooth object surfaces, and smaller windows for surfaces with precipitous depth variations.

Processing defocused images with windows introduces the *window border effect* (Subbarao, 1988). This effect is characterized by the "spread-out" of intensities from neighbouring regions into the window due to image blurring. This "spread-out" is uneven: for a window on the image-plane close to the exact focus position, the intensities are little spread into the window because blurring is little; for a window away from this position the spread-out is more because of more blurring. Therefore, in a sense the window-to-window correspondence dose not hold precisely. This phenomenon always poses a problem in PSF-model based methods (Pentland, 1987; Subbarao, 1988).

However, the problem does not exist in the proposed technique. On one hand, given the window size, this border effect only depends upon the degree of

image blurring. On the other hand, according to the analysis in Section 3.2, the image blurring in one form or another must be a function of t with the shown attribute: symmetry and monotonicity. Therefore, the effect should not change the attribute for $J(t)$, the model for the focal gradient, nor should it affect this technique that is primarily based on the modelling.

3.3.2 Energy in Power Spectrum

Amongst all the known FSC, the energy measure (Subbarao, 1987) that is defined as the normalized *power spectrum* in the window is of particular interest to us. This criterion is well-defined and consistent with others. In addition, it is less vulnerable to noise and, after a conversion from frequency domain to space domain, it is computationally simple and thus less time-consuming. We believe that it is one of the best candidates for representing $J(t)$, the focal gradient function, in this technique.

Referring to Figure 3.1, we begin with a small surface area centred at Q_0, and positioned at t_0. On an image-plane with parameter t, we can always find a window corresponding to that area. The normalized *power spectral density* on that window is then a function of t, and defined as

$$P(f_x, f_y, t) = \frac{|\mathcal{F}(f_x, f_y, t)|^2}{|\mathcal{F}(0, 0, t)|^2} \qquad (3\text{-}8)$$

where $\mathcal{F}(f_x f_y t)$ is the Fourier transform of $f(x,y,t)$, the intensity distribution inside the window on the image-plane associated with t:

$$\mathcal{F}(f_x, f_y, t) = \iint_A f(x, y, t) \cdot \exp[-2\pi j(xf_x + yf_y)]\,dxdy \qquad (3\text{-}9)$$

where A is the area of the window.

From Eq.(3-5), defocus in the spatial frequency domain is primarily characterized by the attenuation of (high) frequencies. Therefore, $P(f_x, f_y, t)$ at a fixed frequency pair (f_{x0}, f_{y0}) would provide a reasonable FSC. In practice, however, selecting (f_{x0}, f_{y0}) can be subjective since the intensity distribution on the target surface is often analytically unknown. Even if the "optimal" (f_{x0}, f_{y0}) can be selected, it is always task-dependent. Moreover, the criterion is vulnerable to noise in the frequency domain.

Indeed, a more robust FSC is the integration of $P(f_x, f_y, t)$ over the entire frequency space, which yields the energy in terms of power spectrum:

$$\mathcal{E}(t) = \iint_B P(f_x, f_y, t)\,df_x df_y \qquad (3\text{-}10)$$

where B is an area in the spatial frequency domain.

Theoretically, area B stretches from $-\infty$ to $+\infty$ in both orthogonal directions in the domain. In practice, it is often an area of finite size depending upon how the image is sampled.

3.3.3 Energy and Grey-Level Variance

$\mathcal{E}(t)$ is not dependent on a specific frequency component and the effect of noise is much reduced. However, $\mathcal{E}(t)$ in the form of Eq.(3-10) involves a large

amount of computation, and special hardware is often required for FFT operations.

In fact, $\mathscr{E}(t)$ can be as well described in the space domain through Parseval's Theorem:

$$\iint_B |\mathscr{F}(f_x, f_y, t)|^2 df_x df_y = \iint_A |f(x,y,t)|^2 dx dy \qquad (3\text{-}11)$$

The *Grey-level variance* (Jarvis, 1976; Krotkov, 1987), which is defined as

$$V(t) = \frac{1}{A}\iint_A [f(x,y,t) - f_m(t)]^2 dx dy \qquad (3\text{-}12)$$

where

$$f_m(t) = \frac{1}{A}\iint_A f(x,y,t)\, dx dy \qquad (3\text{-}13)$$

is directly related to $\mathscr{E}(t)$. Applying Eq.(3-11) to both Eq.(3-10) and (3-12), we obtain

$$\mathscr{E}(t) = \frac{1}{A}\left[1 + \frac{V(t)}{f_m^2(t)}\right] \qquad (3\text{-}14)$$

For a $W \times W$ square window on a digitized image, V(t) can be calculated from

$$V = \frac{1}{W^2}\sum_{x=1}^{W}\sum_{y=1}^{W} |f(x,y) - f_m|^2$$
$$f_m = \frac{1}{W^2}\sum_{x=1}^{W}\sum_{y=1}^{W} f(x,y) \qquad (3\text{-}15)$$

3.3.4 Fast Grey-Level Variance Calculation

A simple "running" scheme can be employed to speed up the grey-level variance calculation. In extracting distances for all or selected points on the normalized images, the $W \times W$ window moves, in either orthogonal direction, L columns from one selected point to the next, where L is the sampling interval on the images. In each transition, the content in the window is updated by throwing away $W \times L$ pixels and adding $W \times L$ new pixels. The remaining $W \times (W-2L)$ pixels are unchanged.

To implement the fast scheme, Eq.(3-15) is rewritten as:

$$V = (f^2)_m - f_m^2$$
$$(f^2)_m = \frac{1}{W^2} \sum_{x=1}^{W} \sum_{y=1}^{W} f^2(x,y) \qquad (3\text{-}16)$$

and f_m and $(f^2)_m$ in the equation are updated through

$$f_m' = f_m + \frac{1}{W^2} \sum_{1}^{W} \sum_{1}^{L} [f(x_a, y_a) - f(x_b, y_b)]$$
$$(f^2)_m' = (f^2)_m + \frac{1}{W^2} \sum_{1}^{W} \sum_{1}^{L} [f^2(x_a, y_a) - f^2(x_b, y_b)] \qquad (3\text{-}17)$$

where f_m' and $(f^2)_m$' are, respectively, the values of f_m and $(f^2)_m$ after the transition; (x_b, y_b) refer to the image points to be abandoned and (x_a, y_a) those to be added.

Under this scheme, the average time(s) each pixel is processed is only about $2W/L$ ($W \geq 2L$) as compared to W^2/L^2 without employing it.

3.4 Algorithms

3.4.1 Focal Gradient Function Model

To estimate the distance (in terms of t_0), the analytical form of J(t) must be assumed. The only knowledge about J(t) is, for the moment, that it is symmetrical and monotonic and thus takes a global extreme at t_0, its geometrical centre (Section 3.2). Obviously, an assumed model is theoretically valid if it is based on, and only on these shown properties of J(t).

Taylor Expansion of J(t) at t_0 provides such a model. Because of its symmetry, only terms to the even power are left in the expansion:

$$J(t-t_0) = a+b(t-t_0)^2+c(t-t_0)^4+..... \qquad (3-18)$$

where a, b, c, ..., are constants.

The monotonicity of J(t) ensures its global extreme at t_0. The analysis shows that t_0 can be analytically estimated from J(t-t_0) in Eq.(3-18) that includes terms up to the fourth power. Two algorithms are then developed that are based on, respectively, the quadratic and the quadruple approximations for J(t-t_0).

In the following algorithms, normalized inverse distances are used for mathematical conciseness. Suppose n images are required in one of the algorithms. The normalized inverse distance is defined as

$$t = \frac{t}{T} - 1 \qquad where \qquad T = \frac{1}{n}\sum_{i=1}^{n} t_i \qquad (3-19)$$

where T is the mean of n known inverse distances, each associated with an acquired image.

3.4.2 Algorithm 1: First Order Approximation

In this algorithm, $J(t\text{-}t_0)$ is only taken to its quadratic term. Three images (n=3) with known t_i are required:

$$J_i \approx a + b(t_i - t_0)^2 \qquad i = 1, 2, 3 \qquad (3\text{-}20)$$

where J_i (i = 1, 2, 3) represents the value of $J(t_i\text{-}t_0)$ obtained from the ith image.

Simple manipulation of Eq.(3-20) yields

$$t_0 = \frac{J_{23}(t_1^2 - t_2^2) - J_{12}(t_2^2 - t_3^2)}{2[J_{23}(t_1 - t_2) - J_{12}(t_2 - t_3)]} \qquad (3\text{-}21)$$

where

$$J_{ij} = J_i - J_j \qquad i = 1, 2, \ldots, n \qquad (3\text{-}22)$$

We can always calibrate the system to satisfy

$$t_1 - t_2 = t_2 - t_3 \qquad (3\text{-}23)$$

Substituting Eq.(3-23) into Eq.(3-21), we have

$$t_0 = t_1 \cdot \frac{J_{13}}{2(J_{23} - J_{12})} \qquad (3\text{-}24)$$

3.4.3 Algorithm 2: Second Order Approximation

One more image is required (n=4) in this algorithm:

$$J_i \approx a + b(t_i - t_0)^2 + c(t_i - t_0)^4 \qquad i = 1, 2, 3, 4 \qquad (3\text{-}25)$$

Manipulation of Eq.(3-25) yields

$$C_3 t_0^2 + C_2 t_0 + C_1 + \frac{C_0}{t_0} = 0 \qquad (3\text{-}26)$$

Mathematically, we can obtain the analytical solution of t_0 from Eq.(3-26) though it is tedious. However, the analysis shows that $C_0 \equiv 0$, if, and only if

$$t_2 - t_1 = t_4 - t_3 \qquad (3\text{-}27)$$

and Eq.(3-26) is reduced to a quadratic. As in Algorithm 1, the relationship in Eq.(3-27) is ensured through the calibration.

Simplifying Eq.(3-26), we obtain

$$\left(\frac{J_{14}}{t_1} - \frac{J_{23}}{t_2}\right) t_0^2 + (J_{12} - J_{34})\, t_0 + \frac{1}{4}\left(t_1^2 - t_2^2\right)\left(\frac{J_{14}}{t_1} + \frac{J_{23}}{t_2}\right) = 0 \qquad (3\text{-}28)$$

Solving Eq.(3-28) yields

$$t_0 = \frac{J_{34} - J_{12} \pm \sqrt{\left(\frac{t_2}{t_1} J_{14} - \frac{t_1}{t_2} J_{23}\right)^2 - 4 J_{12} J_{34}}}{2\left(\frac{J_{14}}{t_1} - \frac{J_{23}}{t_2}\right)} \qquad (3\text{-}29)$$

where "±" is determined from *a priori* knowledge about J(t): what extreme is assumed at t_0. The energy in Eq.(3-10), for example, assumes a global maximum at t_0. Simple analysis shows that $S(t_0) < 0$ for J(t) with the maximum, and $S(t_0) > 0$ for J(t) having the minimum, where

$$S(t_0) = \frac{\frac{J_{14}}{t_1}\left(t_0^2 + t_2^2\right) - \frac{J_{23}}{t_2}\left(t_0^2 + t_1^2\right)}{t_0 \cdot \left(t_1^2 - t_2^2\right)} \qquad (3\text{-}30)$$

If C_3 in Eq.(3-26) happens to be zero, t_0 becomes

$$t_0 = \frac{J_{14}}{2(J_{34}-J_{12})} \cdot \frac{t_1^2 - t_2^2}{t_1} \qquad (3\text{-}31)$$

3.4.4 Other Algorithms

The inverse of the function values in Figure 3.3 yields another J(t) curve as shown in Figure 3.4. Assuming J(t) to approximate a hyperbola:

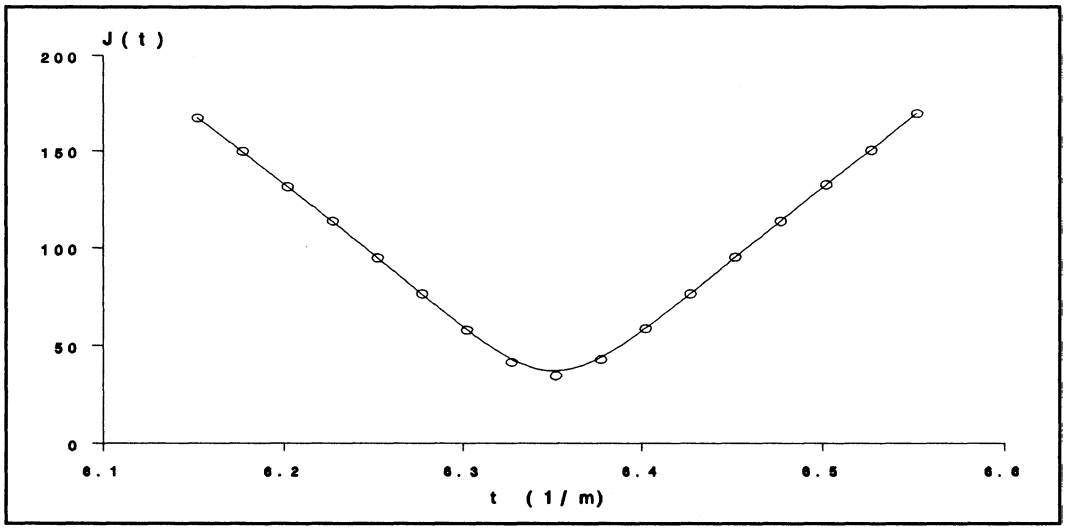


Figure 3.4 $1/\mathscr{E}(t)$ as Focal Gradient Function.

$$J(t-t_0) \approx a+\sqrt{b+c(t-t_0)^2} \qquad (3\text{-}32)$$

leads to another analytical model that may be applied to the inverse energy represented J(t). A simple algorithm can then be developed that requires four (n=4) images. We have

$$J_i = a+\sqrt{b+c(t_i-t_0)^2} \qquad i=1,2,3,4 \qquad (3\text{-}33)$$

On condition that Eq.(3-27) is satisfied, we arrive at from Eq.(3-33)

$$t_0 = \frac{t_1+t_2}{2} \cdot \frac{(J_3+J_4-J_0)J_{34}+(J_1+J_2-J_0)J_{12}}{(J_3+J_4-J_0)J_{34}-(J_1+J_2-J_0)J_{12}} \qquad (3-34)$$

where J_{ij}, $i,j=1,2,3,4$, is defined in Eq.(3-22), and

$$J_0 = \frac{t_1(J_2^2-J_3^2)-t_2(J_1^2-J_4^2)}{t_1(J_2-J_3)-t_2(J_1-J_4)} \qquad (3-35)$$

In fact, it is possible to design different $J(t)$ models based on observations from different FSC employed. The hyperbola model is one example of such approaches. For the moment, except for two universal algorithms in Sections 3.4.2 and 3.4.3, we are unable to give a $J(t)$ model from a specific FSC that is both analytically valid and in good keeping with experimental observation.

CHAPTER 4

IMPLEMENTATION AND TESTING

4.0 Introduction

In this chapter, we demonstrate how the technique is applied to 3-D depth measurement with a computer interfaced imaging system. After an introduction to the system and the related digital image processing facilities, we discuss how the algorithms are realized under given hardware and software environments. Then we attempt to experimentally evaluate its performance under various conditions and provide test results for some real objects.

4.1 System Description

The system consists of an optical imaging device with position-fixed entrance pupil and an image detection and storing device. Certain optical parameter(s) of the system must also be adjustable and recordable. Imaging processing is accomplished by a digital computer, interfaced with the image digitizer for fast image manipulation.

4.1.1 Optical Set-Up

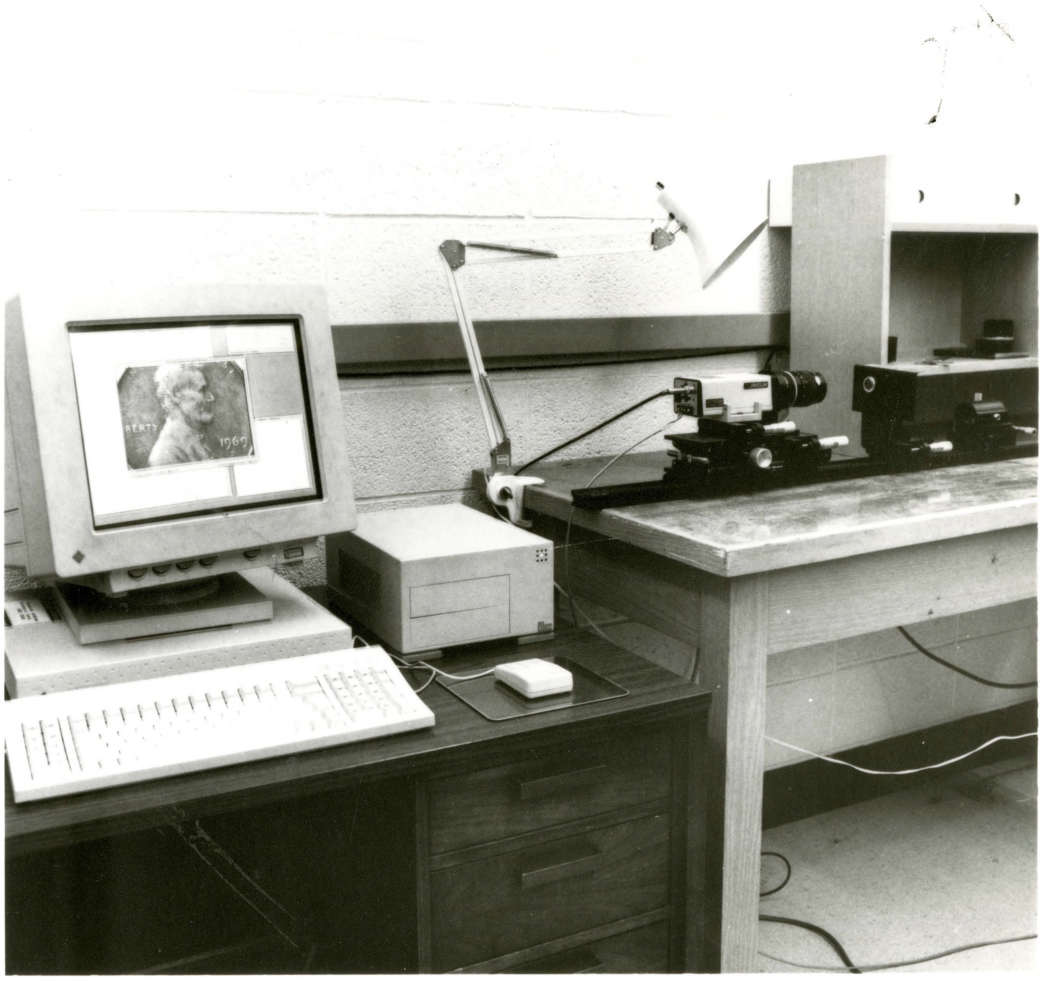


Figure 4.1 Implementing System.

Figure 4.1 shows the implementation of the system used in our experiment. The system includes a Javelin® Ultrichip™ JE-7442 CCD camera and a Nikon Micro-Nikkor® 55mm F/2.8 lens for high quality imaging. The lens is specially designed to produce a flat field image with low distortion. The entrance pupil of the lens is located 14.36mm behind the first optical surface. The Ultrichip camera has better than 52 dB signal-to-noise level. By properly setting the internal control switches, the operation mode of the camera is selected so that linear, natural contour, and low-noise pictures are obtained. The devices, including a bracket holding targets, are all mounted on movable carriers and stages sit on a two-meter optical rail. The translational movements of the devices are controlled by micrometers fixed to the stages and recorded from readings on the micrometer of 2 micron sensitivity. The alignment of the devices along the rail (optical axis) is mechanically ensured and the translational position of the optical axis can be located optically.

In this experiment, all the devices are fixed after the necessary adjustment except for the CCD camera (or the image-plane position), whose translation along the optical axis represents the only allowed change of camera parameters, required to obtain different focus conditions (Section 3.1).

4.1.2 Image Processing Facilities

A Sun® SPARCstation 2 using the Unix operating system is connected with the imaging system. Images from CCD array are grabbed into frame buffers using

the VideoPix® facility. Images are captured at a rate of 30 frames per second and displayed in the window at about 4 frames per second under Preview Mode. The full image (window) size from the VideoPix is 640×480, and the images are quantized and stored in 8-Bit (256) grey-scale TIFF format.

SunIP image processing tools in Sunvision® are used to display and process intermediate 2-D and 3-D data. The SunART advanced rendering tool and SunGV interactive 3-D viewer in the Sunvision package are also used to produce high-quality 3-D visual images from the acquired range images. Figure 1.2(c) in Chapter 1 is an example of such 3-D images.

4.2 Practical Considerations

4.2.1 Image Correspondence

(1) Optical Axis

According to the analysis in Section 3.1.2, the position of the optical axis needs to be first located on the images. In this experiment, the imaging lens is fixed on the optical rail, and within an axial range the optical axis can be considered to be parallel to the optical rail. In principle, two known feature points on the images are required to locate the optical axis of the fixed imaging system. Suppose two $M \times N$ images in Figure 4.2 are taken, for the same object, at two different image-plane positions. O on the images represents the location of the optical axis. S_1 and S_2 are

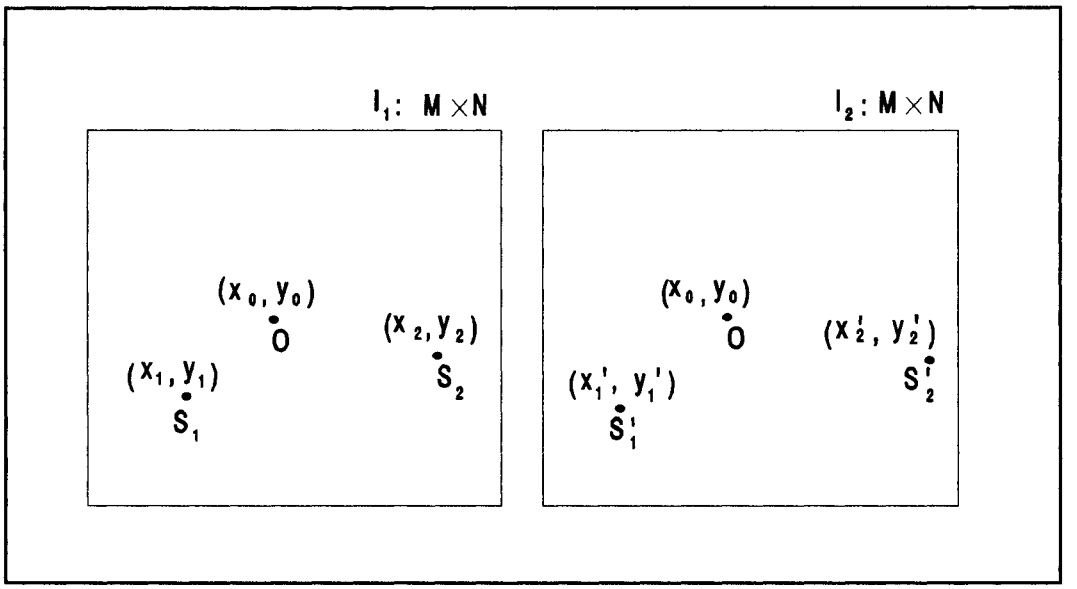


Figure 4.2 Locating Optical Axis.

two feature points on image I_1; S_1' and S_2' are their correspondents on I_2. We have:

$$\frac{x_1-x_0}{x_1'-x_0} = \frac{x_2-x_0}{x_2'-x_0} \quad and \quad \frac{y_1-y_0}{y_1'-x_0} = \frac{y_2-y_0}{y_2'-y_0} \qquad (4-1)$$

or

$$x_0 = \frac{x_1'x_2-x_1x_2'}{(x_1'-x_1)-(x_2'-x_2)}$$
$$y_0 = \frac{y_1'y_2-y_1y_2'}{(y_1'-y_1)-(y_2'-y_2)} \qquad (4-2)$$

Using only two feature points to estimate the optical axis location is erroneous. This is not only because the image resolution is limited by the pixel size, but also because locating the blurred features is often subject to intolerable errors. In this experiment, the following strategy is adopted that involves a large number of feature points:

1) To estimate x_0, a black stripe against bright (white) background, parallel to Y-axis, is used as the target.

2) Two edges of the stripe are taken as the feature lines (integration of N feature points each along the lines);

3) The edge locations (x_1 and x_2) are estimated from the half-intensity points on the blurred edges, in which way the error for locating x_1 and x_2 is significantly reduced. For symmetrical PSF, the error can be practically zero.

4) In addition, taking more than two images and using stripes with different widths, all increase the abundance of valid feature points and thus improve the accuracy of estimating x_0.

5) To locate y_0, the target is turned 90 degrees, and the procedure repeated.

In this experiment, (x_0,y_0) is located to an accuracy of 1~2 pixels within the axial range of about 5mm in object-space. This matching error is neglectable.

(2) Scale Normalization

The coordinates normalization could be carried out by directly zooming the image by a factor as in Eq.(3-2). Depending on which interpolating mechanism is used, nearest neighbour, bilinear, or adaptive, zoomed images are usually different in grey-scale distributions. In other words, zooming a digital image introduces, to a degree, distortion in intensity distribution, especially with the presence of background noise. In this experiment, the normalization is performed without physically zooming the images. As shown in Figure 4.3, the location of point S_1 on the ith image and its

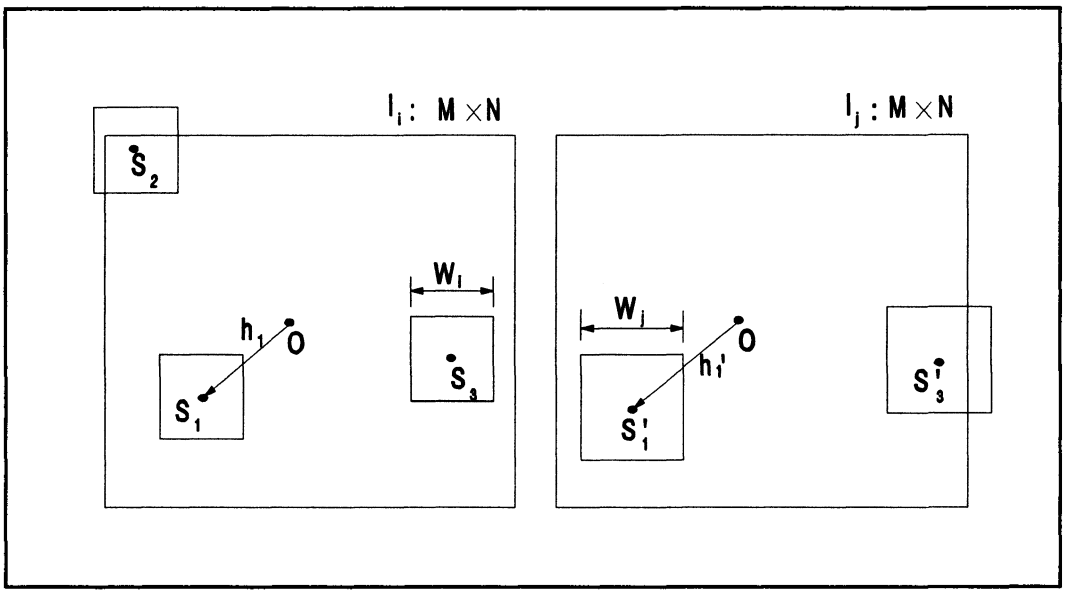


Figure 4.3 Scale Normalization.

correspondent S_1' on the jth image are related through:

$$h_1 = \frac{R_i}{R_j} h_1' \qquad (4-3)$$

and a $W_i \times W_i$ window on I_i corresponds to a $W_j \times W_j$ window on I_j by:

$$W_i = \frac{R_i}{R_j} W_j \qquad i,j = 1,2,\ldots,n \qquad (4-4)$$

where $\boldsymbol{h}_1$ and $\boldsymbol{h}_1$' represent the height vectors for S_1 and S_1' respectively; n is the number of images required in the algorithm. O in the images refers to the optical axis. Applying weights to the pixels, we are able to locate image points at fractional pixel positions.

Image points near image corners and edges require extra attention. The window $W_i \times W_i$ for S_2, for example, exceeds the image border. The window for S_3 is within the image region, but its match on I_j, the $W_j \times W_j$ window for S_3', is not. For

those points, smaller window sizes are needed to ensure the full window-to-window correspondence within the image region.

Note that under this normalization strategy, updating grey-level variance on the images is a bit more complex than what is described in Eq.(3-17). The calculation of focus sharpness criteria (FSC) from the window also needs slight modification. Consider $\mathscr{E}(t)$, the energy measure for example. Suppose that $f(x,y)$ represents the normalized intensity distribution in the window and A the normalized window size. The intensity distribution in the original (non-normalized) window is then $a \cdot f(x/R, y/R)$, and the window area $A \cdot R^2$, where a is a constant and R is (refer to Eq.3-2) the normalization factor. $\mathscr{E}(t)$ in Eq.(3-14) represents a calculation from the scale normalized window, and let $\mathscr{E}_R(t)$ be a calculation from the original window. Similar to the derivation of Eq.(3-14), we obtain

$$\mathscr{E}_R(t) = \frac{1}{R^2}\mathscr{E}(t) \qquad (4-5)$$

In practice, $\mathscr{E}_R(t)$ can be modified by simply multiplying itself with the window area (which is proportional to R^2):

$$\mathscr{E}(t) = (AR^2) \cdot \mathscr{E}_R(t) \qquad (4-6)$$

in which way, referring to Eq.(3-14), $\mathscr{E}(t)$ is also "regulated" to have unit base value regardless of window area.

Subtracting 1 from Eq.(4-6), we can even further regulate the $\mathscr{E}(t)$ value to have zero base. In fact, such a regulated form of $\mathscr{E}(t)$ is equivalent to the

normalized variance in the window. In this experiment, unless otherwise specified, $\mathscr{E}(t)$ is always calculated and referred to this way, and energy and grey-level variance are used interchangeably where applicable. The energy function in Figure 3.3, for example, is presented in this regulated form.

4.2.2 Camera Response and Noise

For fixed camera parameters and under constant illumination, two images obtained at different times have different intensity distributions. This was also observed by Krotkov (1987), and termed "temporal variations". The variations are not due to the varying incandescent illumination from time to time since they appear to be random over the image region even under point source illumination. We attribute the temporal variations primarily to the thermal noise on the CCD array in the camera.

Subtracting two images taken for the same object at different times yields an *image of temporal variations*. Statistically, the variance from this image is twice as much as the variance of the actual noise. Applying illuminations of different radiant intensities to a plain object, we are able to extract, under our experimental condition, the relationship between the noise and the average grey-level on the images. Curve I in Figure 4.4 shows the result. According to this curve, the noise level is about constant over a range, and the average signal-to-noise level is about 40 dB.

Acquired in a similar way, curve II in the figure shows the camera response in terms of grey-level *vs* relative image intensity. In this experiment, all the

concerned images are taken within a grey-level range of about 50~220 so that both a linear response and a near constant noise level hold over the range.

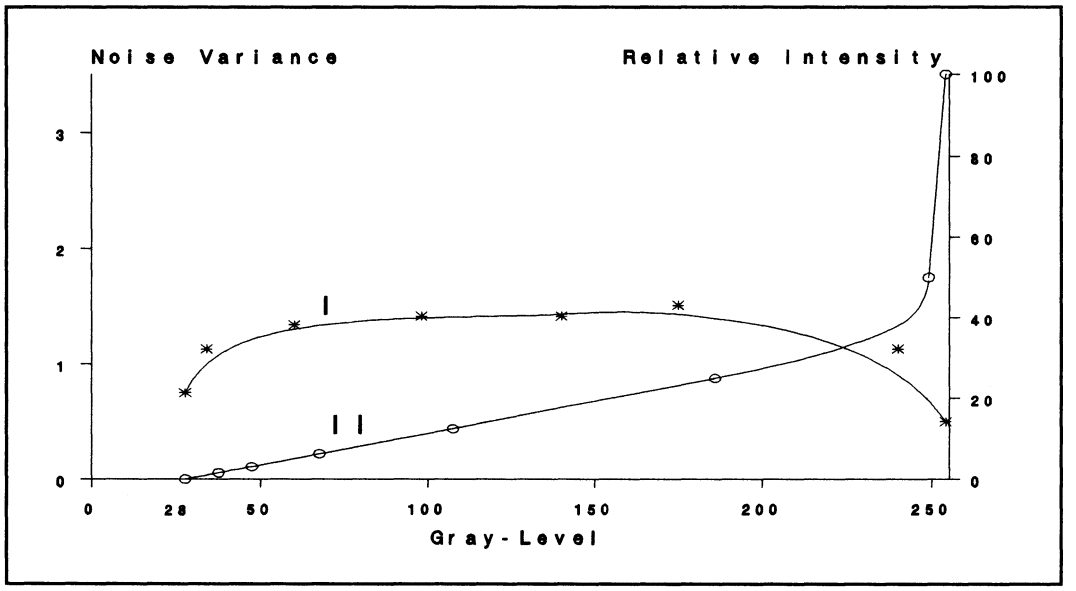


Figure 4.4 Camera Response and Noise Level.

The presence of thermal noise also results in the CCD camera having non-zero response without any incoming light. It is observed that with an average of 28 grey-level, this background response is unevenly distributed over the CCD plane. Simple analysis shows that the symmetry of $\mathscr{E}(t)$ is affected by this response, especially when images are acquired under low illumination level. In this experiment, this non-zero and spatially varying response is removed by subtracting the images and an image, obtained once for all by averaging a number of images taken with no incoming light.

Nevertheless, the effect of thermal noise on this technique is minimal. This is not only because the noise is comparatively small and no noise-vulnerable differentiation operation is involved, but also because the noise can be statistically

eliminated through the image subtraction and the window operation.

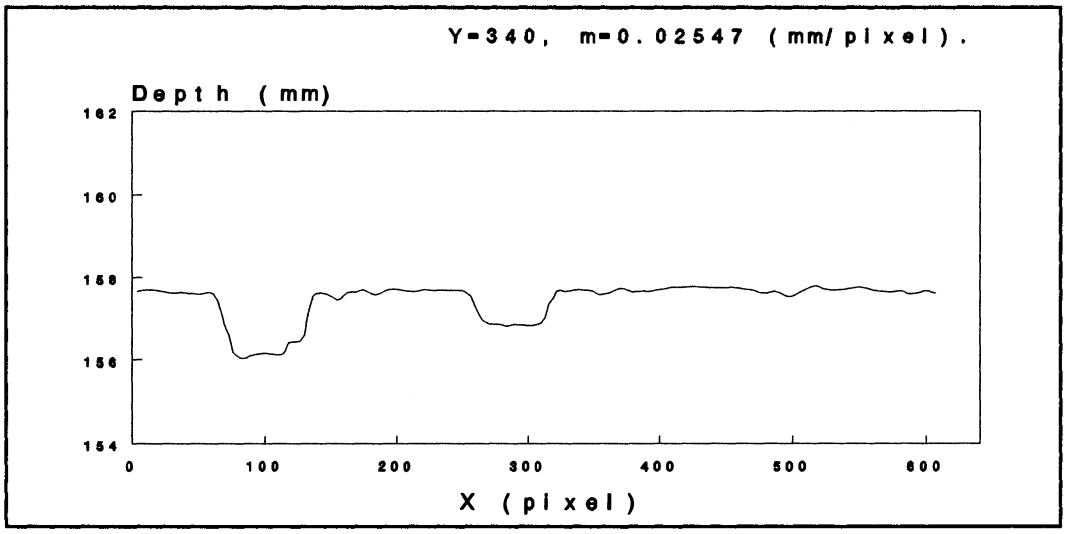
4.2.3 Surface Directional Reflection

It is no doubt that this technique fails for textureless surface areas for the lack of spatial frequency content. Subbarao (1989) discussed the problem in general and provided a solution that introduces "texture" by controlled illumination.

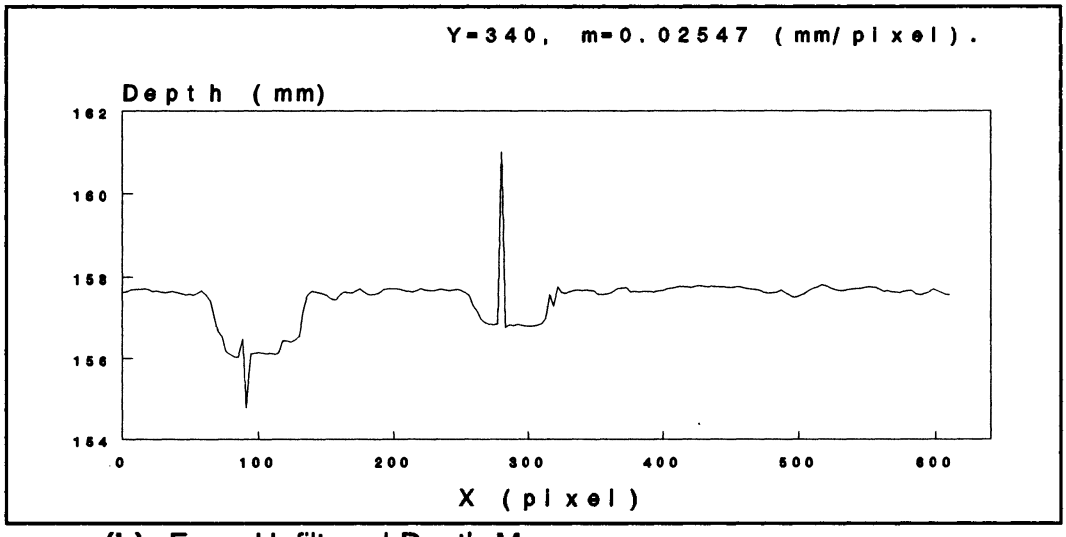
Most machine parts are not Lambertian reflectors either. The sporadic directional reflection from the surface often results in saturated brightness on the images and causes significant distortion in distance estimation. The glare effect may remain for surface features such as edges and corners even if the surface is physically or chemically processed. A polarizing filter may be used to reduce the effect. However, we observe only limited improvement using the filter, and the improvement is often countervailed by the reduced incoming light from non-glaring surface areas.

Median filtering is an effective way to remove scattered noise (Nayar, 1992), especially when the noise amplitude is relatively high. A fast 2-D median algorithm (Huang, 1979) is employed in this experiment to reduce the noise on depth maps, caused primarily by the directional reflection from surface edges and corners. The depth map in Figure 1.2(a), for example, is the improved result after 5×5 median filtering. Figure 4.5(a) shows one cross-section of the depth map; (b) gives the corresponding cross-section from the unfiltered depth map. The improvement is distinct, and, as long as the filtering window is small compared with the window for distance estimation, the surface "smoothing" effect brought about by the filtering is

negligible.



(a) From Median Filtered Depth Map.



(b) From Unfiltered Depth Map.

Figure 4.5 Cross-Sections of Depth Maps.

4.3 Calibration Scheme

The aim and the necessity for camera calibration in this technique are stated in Section 2.2.3. In principle, the relationship between t_i, $i=1,2, ..., n$, and a camera

setting is determined, in the calibration, by the fact that under this camera setting, the image taken from t_i is exactly in focus (where $J(t_i)$ takes its extreme value). The number n here refers to the number of t_i (object-planes) calibrated, which must be larger than the number of images required in a specific algorithm. The system can be calibrated either by searching for camera settings that are associated, respectively, with given t_i, or by looking for object-distances $l_i = 1/t_i$ which correspond, respectively, to given (the ith) camera settings. Since Eq.(3-23) and (3-27) must be satisfied to simplify the algorithms, the first option is, at least for Algorithm 2, the only choice.

4.3.1 Searching for Camera Settings

In general, the first option can be described (refer to Figure 3.1) as follows:

1) Position a target at P_i with known t_i and known height, say Q_iO_i;

2) Accommodate camera parameter(s) and find the sharpest image;

3) Record $Q_i'O_i'$ (or $m_i = Q_i'O_i'/Q_iO_i$) and the camera setting;

4) Repeat 1) ~ 3) n times and the system is ready to go.

An object of contrast step-edges, like one used in locating the optical axis, may serve as the target. In practice, however, to accurately locate, along the optical axis, where the sharpest edge occurs is found severely affected by both thermal and discretization noises. Moreover, because of the existence of astigmatic difference in the system, the axial location of the sharpest edges is affected by edge orientations.

Worst of all, it is noticed that, depending on which edge finding algorithm is

used, the sharpest edge does not always correspond with no sensible difference to the maximum energy — although all focus sharpness criteria (FSC), including the energy and the edge measures, should in principle agree with each other. In consideration of these factors, the following two-phase calibration scheme is adopted:

Phase 1. Acquiring magnification distribution using edges:

a) Position a stripe target of known width perpendicular to the optical axis at P_i with known t_i (refer to Figure 3.1). The stripe is placed parallel to X direction.

b) Move the image-plane about P_i' along the optical axis, obtaining a series of images, the number of which should be large enough to have a good estimate of m_{xi} distribution about P_i', $m_{xi}(l')$.

c) For each image, find m_{xi} and record the relative position of the image-plane. Note that edges are not necessarily in focus here and edge finding techniques must be applied (Shirai, 1987).

d) Interpolate the data and obtain $m_{xi}(l')$.

e) Place the target along Y-axis, repeat a)~d), obtaining the ratio of m_x/m_y.

f) Repeat a)~d) n times.

Phase 2. Acquiring the records of camera settings using energy measure:

a) Position a plane target perpendicular to the optical axis at P_i with given t_i. For the best performance, there should be patterns on the target plane that are rich in spatial frequency components. Figure 4.17(b) in the next section shows what the patterns look like in this experiment. This kind of target is

referred to as *standard target*.

b) Move the image-plane about P_i' along the optical axis, obtaining a series of images. Searching strategies may be applied here to minimize the number of images required.

c) Normalize the images based on $m_{xi}(l')$ and m_x/m_y obtained in Phase 1.

d) Calculate the energy $\mathscr{E}(l')$ on the images and search for the maximum.

e) Record the camera setting corresponding to the maximum.

f) Repeat a)~e) n times.

In principle, n can be very large that a great range of depth is covered in object-space. This way, the system gains greater flexibility: the object can be placed either near or far away from the system according to the requirements of field of view and depth resolution.

To calibrate a large number of t_i is a time-consuming and tedious task that is beyond the scope of this experiment. In this thesis, we are mainly concerned about the effectiveness and the accuracy of the new technique rather than its flexibility. At an average stand off distance of 157.42mm from the system, a total of five ($n=5$) object-plane positions are calibrated over a range of 4.96mm. For a specific algorithm, only part (3 or 4) of the calibrated planes are used. The calibration result is presented in Table 4.1 where p_i refers to the record of the camera parameter setting, acquired from the micrometer fixed to the CCD camera.

Because the same FSC (inverse energy) is used in both calibration and measurement procedures, there is no problem of disagreement between different

FSC. Using the same type of light source in both procedures also minimizes the effect of chromatic aberration in the system.

Table 4.1 Calibration Result.

i	l_i (mm)	t_i (1/mm)	p_i	m_{xi} (pixel/mm)	m_{yi}/m_{xi}
1	154.98	0.0064524	13.5~05.8	40.1756	1.00944
2	156.19	0.0064024	13.5~40.8	39.7167	
3	157.42	0.0063524	14.0~25.5	39.2618	
4	158.67	0.0063024	14.5~09.9	38.8107	
5	159.94	0.0062524	14.5~44.0	38.3636	

4.3.2 Calibration Interval

In this calibration, the interval between each neighbouring pair of calibrated object-planes is set for 0.05 1/m and referred to as *calibration interval.* For the convenience of the discussion that follows, we define:

1) the interval between the first and the last calibrated object-planes involved in a algorithm as the *range of implementation*, or *I-range.* In this experiment, the I-range may include 2~4 calibration intervals depending upon which algorithm and how the algorithm is implemented.

2) the range of the central region on the J(t) curve, where a good quadratic or quadruple approximation holds, as *Q-range.* There is no absolute criterion for determining Q-range. It depends upon the requirement for depth resolution, the *depth of field* of the system, and even the form of J(t). The J(t) curves in Figures 3.3 and 3.4, for example, apparently have different Q-ranges given the same criterion for

determining them.

3) the range where linear and low-error-level distance estimates are acquired as the *range of measurement*, or *M-range*.

In a measurement, the J(t) curve moves along the optical axis as the surface depth changes, as does the Q-range. Generally speaking, to obtain a certain M-range, the Q-range, wherever it moves, must cover the I-range. Under such a condition, the sum of I-range and M-range must equal Q-range. In other words, there is a limit for I-range, and the I-range too large may result in intolerable distortion in distance estimation. On the other hand, a small I-range only contains a small section of Q-range, within which the difference between the measured J(t) values may be too little to overcome the measurement noise and the error of low-order approximations within the Q-range.

In short, the range of implementation should be such that a good quadratic or quadruple approximation for J(t) is ensured within certain range of measurement, and the calibration interval can then be selected accordingly.

For instance, as seen from the J(t) curves in Figures 3.3 and 3.4, the interval for this calibration (0.05 1/m) appears, at least in Algorithm 1 (quadratic), a bit larger for the energy represented J(t), but is fairly reasonable for J(t) described by inverse energy (variance). This observation is further examined through a test in Section 4.4.1.

4.4 Experimental Results

The algorithms are realized in the C programming language for a SPARC station under the Unix environment. The algorithms are themselves fairly straightforward. Image pre/post-processing deserves special attention. The weighted and fractional pixel, for example, is used in the image processing procedures to ensure accurate image correspondence and mapping. The system adjustment and the calibration procedures are time-consuming. The programme includes four major subroutines:

1) Image acquisition: grab images into frame buffers; if necessary, average the images to reduce the noise effect.

2) Pre-processing: remove the unevenly distributed noise level on the images; normalize the image scales.

3) Distance estimation: apply the algorithm to all or selected image points; obtain the depth map.

4) Post-processing: median filter the depth map; convert the depth map to Cartesian coordinate system; acquire 3-D plot if necessary.

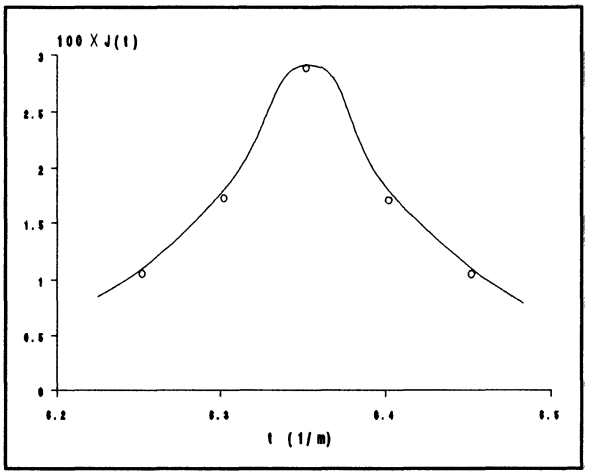
In this experiment, we choose $R_3=1$, so the scale factor is (refer to Eq.3-2):

$$K = m_3 l_3 \qquad (4-7)$$

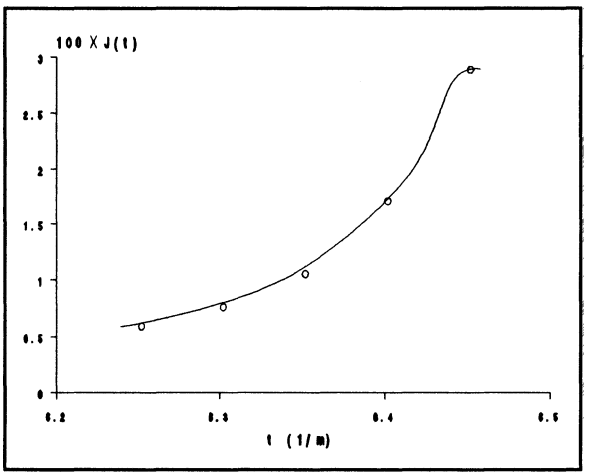
In other words, whenever a $W\times W$ window is applied to the images in a measurement, it really means that only on the image taken with $p_3=14.0\sim25.5$ (corresponding to $l_3=157.42$mm) does the window size equal to $W\times W$.

4.4.1 Tests Using Standard Target

The system and the algorithms are evaluated with the same standard target as is used in the calibration. Over the calibrated range in object-space (154.98mm~159.94mm), the target is placed at 13 different positions. At each position, 5 images are acquired with different calibrated sets of camera parameters. Depending upon the aim of each test, the images are averaged over 1~8 frame(s) taken at different times. To evaluate the effect of aperture on the accuracy in depth estimation, over a dozen images are also taken with different aperture sizes.



(a) l_0 = 157.42 mm.



(b) l_0 = 159.94 mm.

Figure 4.6 Focal Gradient Function Values.

(1) Acquiring J(t) Curves

For each target position, 5 J(t) values can be extracted from the images. An approximate estimate of J(t) curve for that position can then be obtained over the calibrated range. Figure 4.6(a) is such an example when the target is placed at 157.42mm (t_0=0.0063524 1/mm); (b) shows the result when t_0=0.0062524 1/mm. In

these examples, each image is averaged over two frames, and the energy in a 400×400 window is calculated to represent the value of J(t).

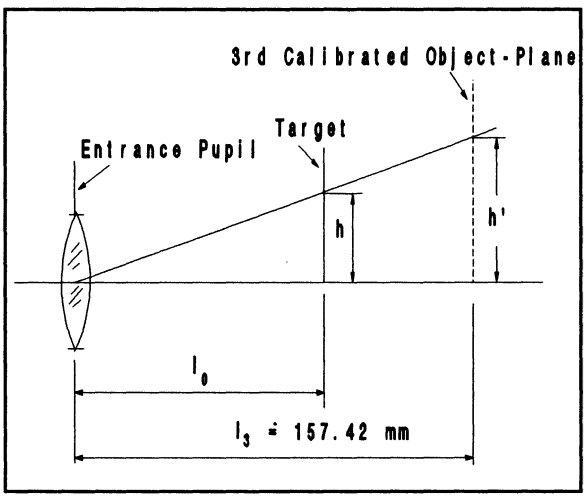


Figure 4.7 Target Positions.

Obviously, each curve only represents a section of J(t). It is not difficult to combine these sectional curves into a J(t) curve over a larger range and with more interpolating points. The only problem is to ensure that the same surface area is processed for all the target positions. As shown in Figure 4.7, given $R_3=1$ and assuming the target moves along the optical axis within the range, we have:

$$W = m_3 h' \qquad h' = \frac{l_3}{l_0} h \qquad (4\text{-}8)$$

or from Eq.(4-7):

$$W = \frac{K \cdot h}{l_0} \qquad (4\text{-}9)$$

where h represents the height of a square surface area to be processed on the target, and h' its projection on the calibrated object-plane; K, the scale factor, is a constant; l_0 represents the target position, and W the window width corresponding to the height of that surface area.

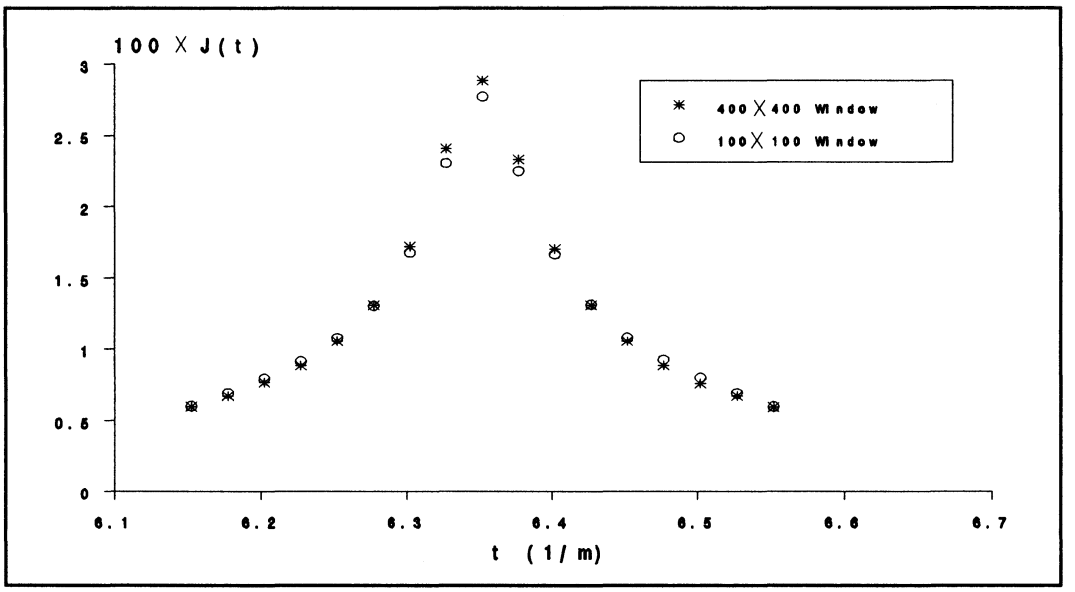


Figure 4.8 $\mathscr{E}(t)$ as Focal Gradient Function.

Therefore, to obtain J(t) curves that come from the same surface area, one has to apply different window sizes while moving the target from one position to another. For example, if we choose W=400 for a measurement, then the actual window size for l_0=l_3 is the same, while that for l_0=l_5=159.94mm must be smaller (about 394).

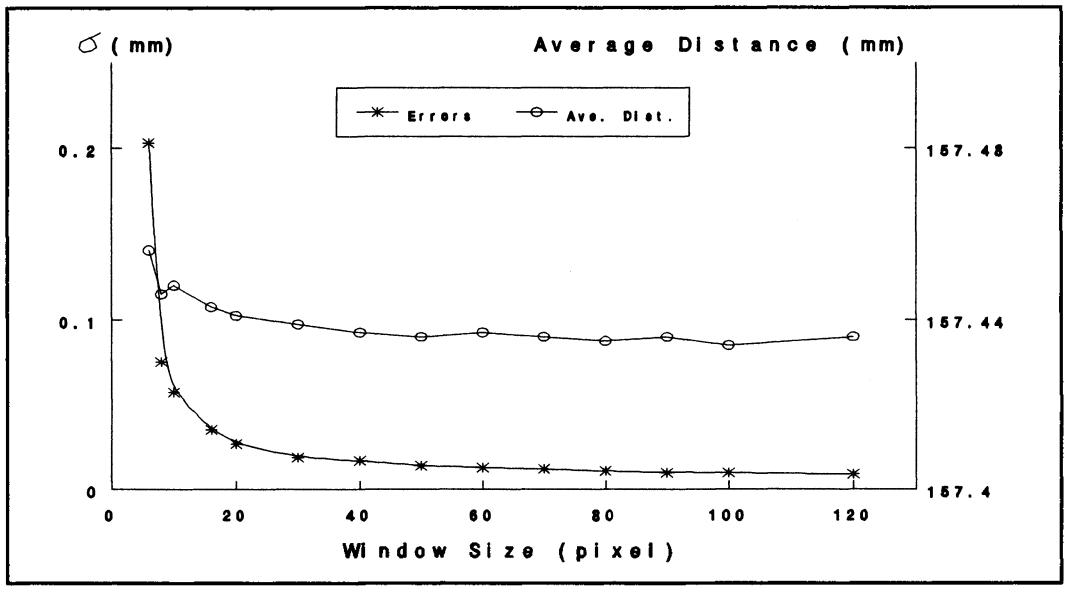
Table 4.2 Values (×100) of Focal Gradient Function.

t (1/m)	6.152	6.177	6.202	6.227	6.252	6.277	6.302	6.327
W=100	0.600	0.689	0.790	0.911	1.072	1.301	1.679	2.307
W=400	0.598	0.668	0.762	0.881	1.054	1.307	1.722	2.408
6.352	6.377	6.402	6.427	6.452	6.477	6.502	6.527	6.552
2.774	2.251	1.667	1.309	1.075	0.922	0.795	0.688	0.595
2.887	2.331	1.704	1.307	1.050	0.882	0.756	0.666	0.590

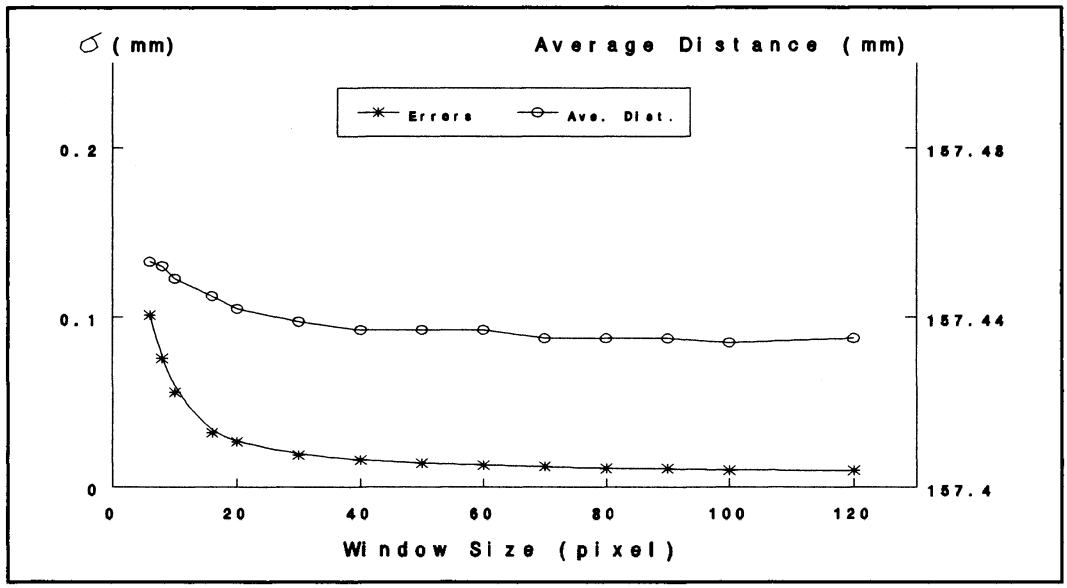
Figure 4.8 and Table 4.2 gives the results of our calculation from two window sizes, 400×400 and 100×100, assuming the target position l_0=l_3. In fact, Figure 3.3

in Chapter 3 is from one of the results (400×400 window).

(2) Effect of Window Size



(a) $W = L$.



(b) $L = 30$.

Figure 4.9 Errors & Window Sizes.

We now examine how the accuracy in depth estimation is affected by the size of windows. The standard target is placed at l_3=157.42mm. The algorithms are tested under different conditions and the results are generally the same, *i.e.*, they are all, in principle, in agreement with the analysis in Section 3.3.1. Two typical results from Algorithm 1 are presented in Figure 4.9. No temporal averaging is performed on the images in this test. In the figure, σ represents the standard (root-mean-square) deviation of distance estimates for a large number of surface points on the target plane.

Statistically, to obtain a satisfactory estimate of distance errors, it is unnecessary to take into account distance estimates from all the surface points. The images involved in obtaining the results in this section are all properly sampled. Sampling interval L in Figure 4.9(a) is equal to the window width, which way, regardless of window size, all the image points are processed just once in a measurement; L in (b) is kept to 30 pixels in both orthogonal directions on the images. In the following test examples in this section, unless otherwise noted, the window size always equals the sampling interval on the images.

The original data from the figure are listed in Table 4.3. Two sets of the result are basically the same, and W=30 (pixel) is the place where the error of distance estimates is about twice as much as the minimal error ($\sigma_{min}\approx$0.01mm in this measurement).

In this and the following two tests, the second, the third, and the fourth calibrated positions in object-space are involved in distance estimation, using

Algorithm 1 and with inverse energy representing J(t). This experimental scheme is expressed, in this thesis, as 2-3-4.I. "I" here stands for Inverse energy, and "E" and "H" are used to denote the Energy measure and Hyperbola algorithm respectively. The "1-2-4-5.E" scheme, for example, indicates that, with the energy representing J(t), Algorithm 2 is applied and the 1st, 2nd, 4th, and 5th calibrated object-planes are involved.

Table 4.3 Distance Estimates, Errors, and Window Sizes.

W	$W = L$		$L = 30$	
	Ave. Dist. (mm)	σ (mm)	Ave. Dist. (mm)	σ (mm)
6	157.456	0.203	157.453	0.101
8	157.446	0.075	157.452	0.076
10	157.448	0.057	157.449	0.056
16	157.443	0.035	157.445	0.032
20	157.441	0.027	157.442	0.027
30	157.439	0.019	157.439	0.019
40	157.437	0.017	157.437	0.016
50	157.436	0.014	157.437	0.014
60	157.437	0.013	157.437	0.013
70	157.436	0.012	157.435	0.012
80	157.435	0.011	157.435	0.011
90	157.436	0.010	157.435	0.011
100	157.434	0.010	157.434	0.010
120	157.436	0.009	157.435	0.010

(3) Time Averaging

Averaging the images taken at different times is a simple way to reduce the

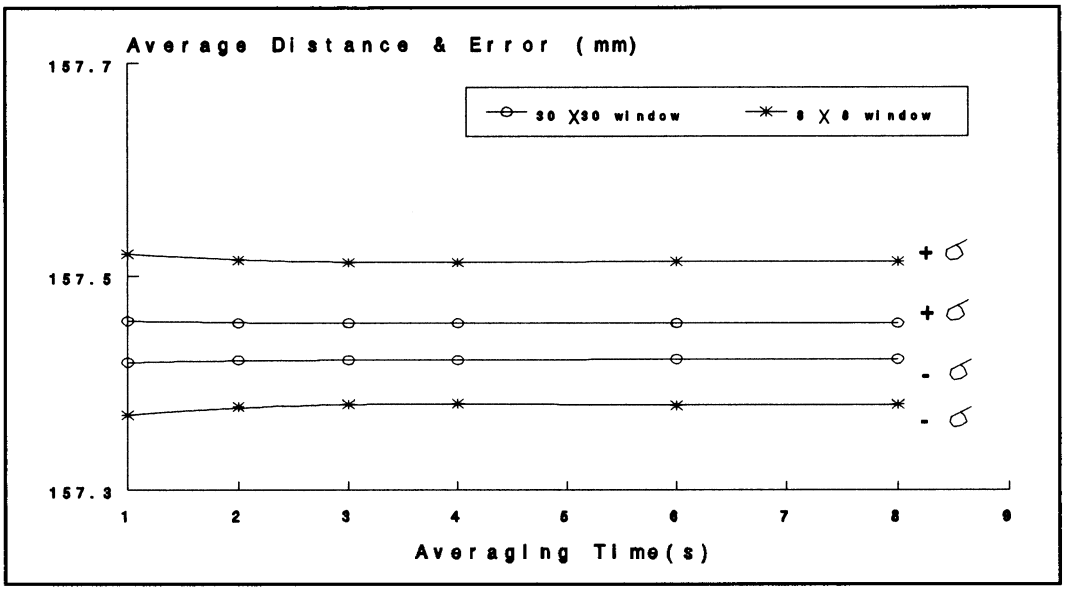


Figure 4.10 Errors and Averaging Time(s).

temporal noise. Two examples are provided in Figure 4.10 and Table 4.4 to demonstrate the effect. The target is placed at a stand-off distance 157.42 mm from the entrance pupil and two window sizes are used in the algorithm (using 2-3-4.I scheme).

Table 4.4 Distance Estimates, Errors, and Averaging Time(s).

Ave. Time(s)	W = 8		W = 30	
	Ave. Dist. (mm)	σ (mm)	Ave. Dist. (mm)	σ (mm)
1	157.446	0.075	157.439	0.019
2	157.447	0.068	157.439	0.017
3	157.447	0.066	157.439	0.017
4	157.447	0.065	157.439	0.017
6	157.447	0.067	157.439	0.016
8	157.447	0.067	157.439	0.017

It is observed that to average the images more than two times brings about little more improvement on the accuracy in distance estimation. For larger windows

(*e.g.* 30×30), without resorting to time averaging, the temporal variations can almost be eliminated only through window operation, which is indeed consistent with the analyses in Sections 3.3.1 and 4.2.2.

In the following tests, each image is averaged twice in the time domain, which, according to this observation, reduces the effect of temporal variations to a negligible level (along with the measure removing the average noise level).

(4) Effect of Aperture Size

The size of aperture plays an important role in determining the *depth of focus* or the error of depth in object-space (a discussion on the topic is in Section 5.1). Figure 4.11 shows, in an example, how the estimated distance is affected by aperture size. In this example, the standard target at 157.42 mm is tested using the 2-3-4.I scheme and with a 30×30 window.

It is observed that the performance is fairly consistent from F/2.8 to F/4. After that, the accuracy drops constantly. Although measures are taken to increase the incoming light as the aperture size decreases, the intensity level on the images under F/11 is still much lower than normal, which is probably the reason that a much larger error is observed under this aperture. Table 4.5 provides the original data from this test.

Table 4.5 Distance Estimates, Errors, and Aperture Sizes.

Aperture	F/2.8	F/4.0	F/5.6	F/8.0	F/11
Ave. Dist. (mm)	157.43	157.42	157.22	157.06	157.07
σ (mm)	0.017	0.021	0.036	0.077	1.67

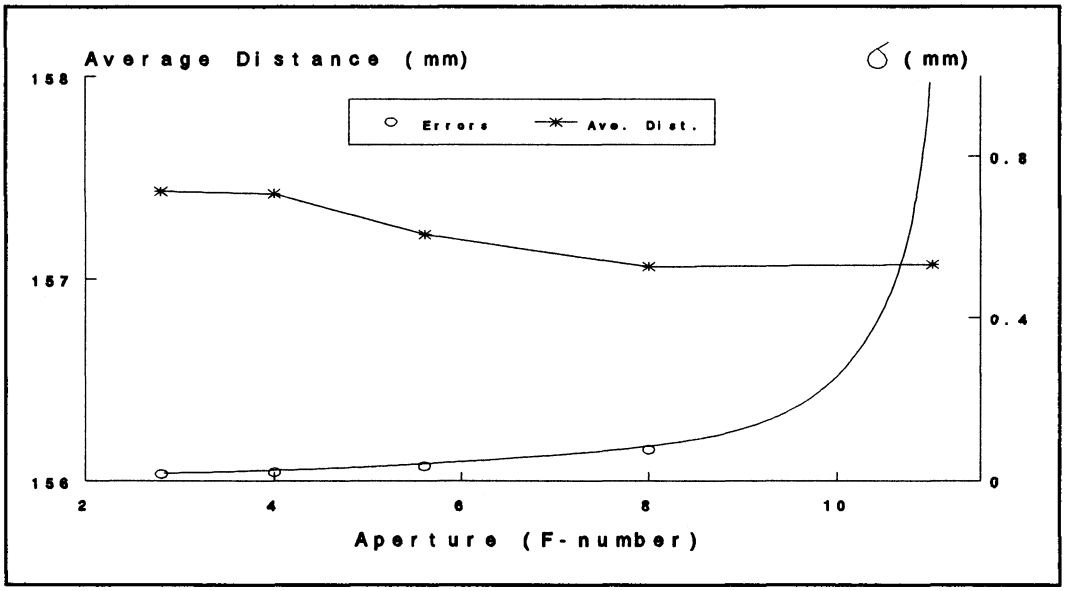


Figure 4.11 Errors and Aperture Sizes.

In this experiment, except for those in this test, all images are taken under F/2.8 aperture to obtain the maximum focus effect (or the minimal depth of focus).

(5) Range of Measurement

This test is intended to examine, in an example, how the error in distance estimation is related to the range of measurement (M-range) and affected by I-range. The *distance deviation* at a target position is defined as:

$$Err. = e \pm \sigma \qquad (4\text{-}10)$$

where σ is the standard deviation of estimated distances; and e denotes the difference between l_0^*, the average of the distance estimates, and l_0, the real distance:

$$e = l_0^* - l_0 \qquad (4\text{-}11)$$

Generally speaking, for a given target position, σ represents the relative (random) error in distance estimation, and e the absolute (systematic) deviation. The value e varies over the range of measurement and its root-mean-square error, σ_e, can be estimated from different target positions within the range.

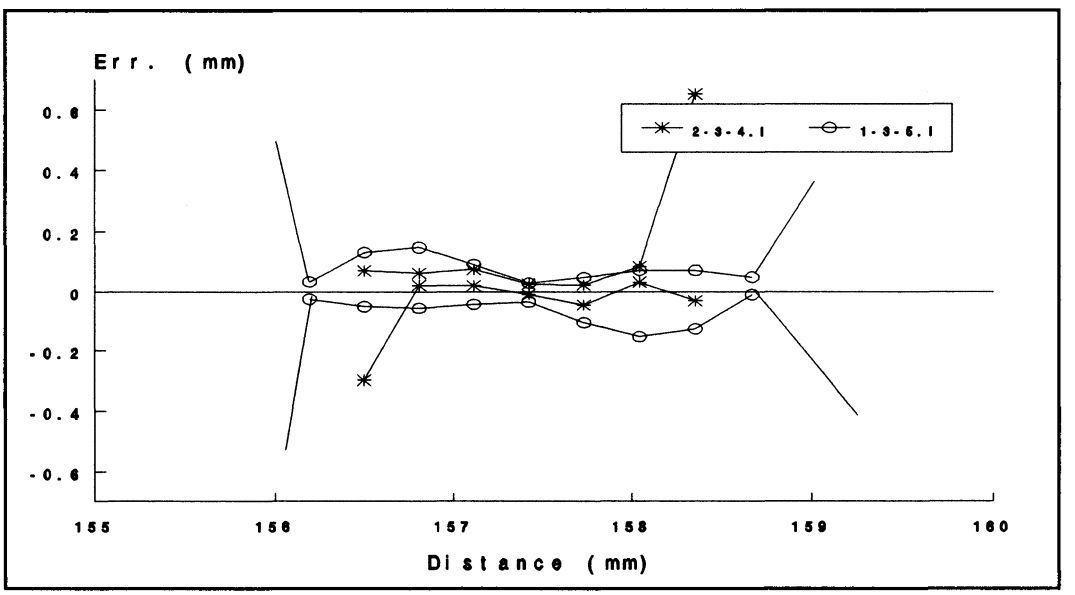


Figure 4.12 Errors and Ranges of Measurement.

The 30×30 window is used in this test and, unless otherwise specified, the same window size is applied to the rest of the tests in this section. Figure 4.12 and Table 4.6 provide the result in this test, where the standard target is tested, under 1-3-5.I and 2-3-4.I schemes, from 13 different positions over the entire range of calibration (154.98mm ~ 159.94mm). The "—" in the table indicates that no meaningful result can be extracted at that distance with the given scheme and window size.

The I-range for the 1-3-5.I scheme doubles that for the 2-3-4.I scheme. It is observed that the M-range for the former is about twice as large as that for the latter

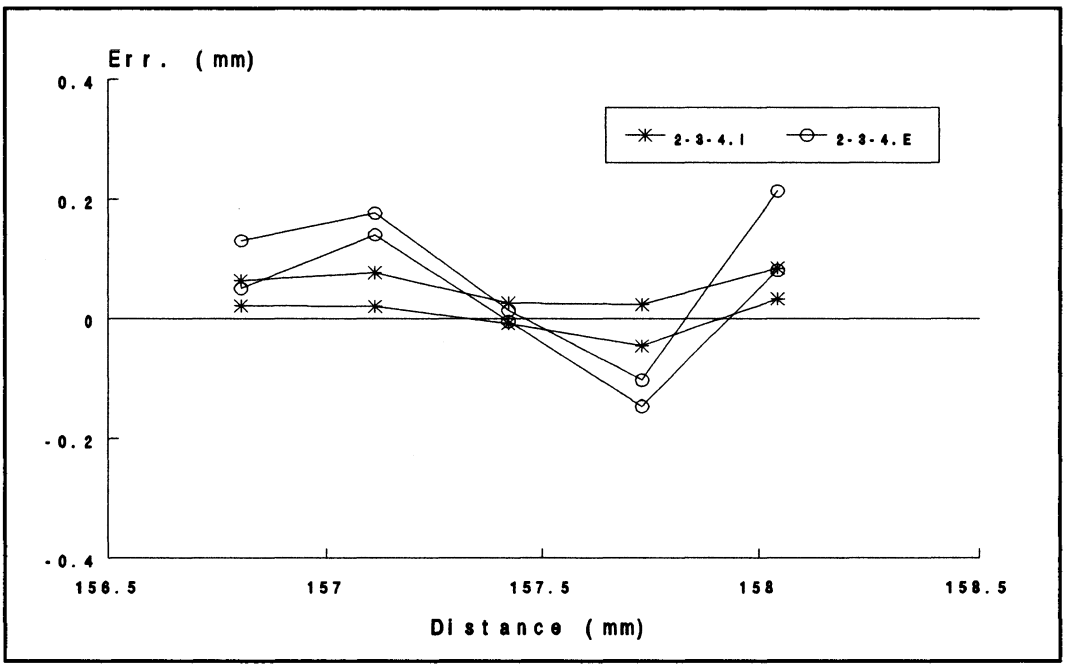
(errors measured within M-ranges are bold-faced in the table). According to the analysis in Section 4.3.2, the Q-range for the 1-3-5.I scheme should also be twice as much as that for the 2-3-4.I. One should not be surprised at the result since a larger average error is observed within the M-range for the 1-3-5.I, and so is a larger Q-range.

Table 4.6 Distance Estimates and Errors: Inverse Energy Measure.

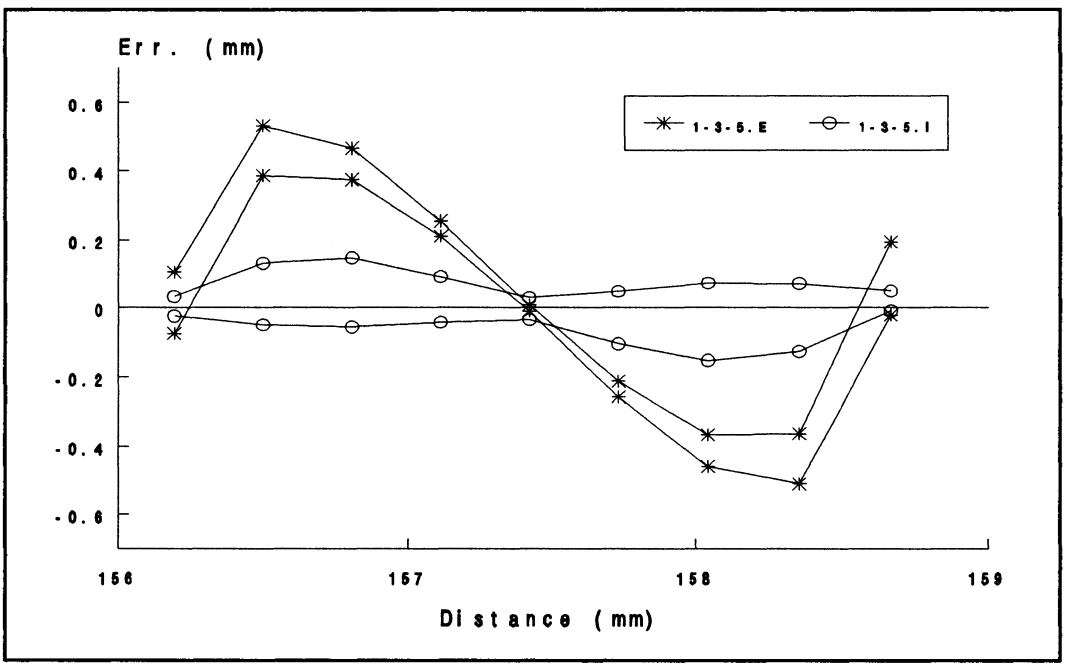
Target Posi. l_0 (mm)	2-3-4.I		1-3-5.I	
	Ave. Dist. l_0^* (mm)	σ (mm)	Ave. Dist. l_0^* (mm)	σ (mm)
154.981	—	—	—	—
155.584	—	—	155.741	7.015
156.191	—	—	156.195	**0.029**
156.497	156.383	0.185	156.538	**0.092**
156.804	156.846	**0.021**	156.850	**0.102**
157.112	157.160	**0.028**	157.136	**0.068**
157.421	157.430	**0.018**	157.417	**0.032**
157.731	157.720	**0.034**	157.702	**0.077**
158.043	158.101	**0.026**	158.003	**0.113**
158.356	158.667	0.342	158.328	**0.099**
158.670	—	—	158.689	**0.029**
159.302	—	—	159.801	2.963
159.939	—	—	—	—

(6) Energy and Inverse Energy Measures

The J(t) represented by inverse energy has been used in testing the effects of window, time-averaging, and aperture. It is noticed from Figures 3.3 and 3.4 that this J(t) form, as compared with the energy represented J(t), appears to be closer to the



(a) 2-3-4.I *vs* 2-3-4.E.



(b) 1-3-5.I & 1-3-5.E.

Figure 4.13 Energy *vs* Inverse Energy Measures.

low-order Taylor approximations (or have a larger Q-range). The intuition is further examined in this test.

Two of the "E" schemes are tested for different target positions. The result is presented in Table 4.7. The results from 2-3-4.I and 2-3-4.E are compared in Figure 4.13(a), and those from 1-3-5.I and 1-3-5.E are in (b). Only data in the M-ranges are shown in the figure. It is observed that the inverse energy is superior to the energy measure in performance. Because of the smaller Q-range, strictly speaking, there is no linear range of measurement (M-range) for either of the "E" schemes.

Table 4.7 Distance Estimates and Errors: Energy Measure.

Target Posi. (mm)	2-3-4.E		1-3-5.E	
	Ave. Dist. (mm)	σ (mm)	Ave. Dist. (mm)	σ (mm)
154.981	159.040	0.581	159.488	0.511
155.584	158.857	0.378	—	—
156.191	160.334	3.056	156.206	**0.091**
156.497	—	—	156.956	**0.072**
156.804	156.894	**0.040**	157.224	**0.045**
157.112	157.271	**0.018**	157.345	**0.022**
157.421	157.425	**0.009**	157.419	**0.009**
157.731	157.606	**0.022**	157.497	**0.022**
158.043	158.190	**0.067**	157.630	**0.047**
158.356	—	—	157.919	**0.074**
158.670	155.219	2.190	158.757	**0.109**
159.302	155.914	0.567	152.767	8.640
159.939	155.798	0.722	155.405	0.603

(7) Algorithms

With inverse energy representing J(t), Algorithm 2 is tested under two slightly different schemes: 1-2-3-4.I and 1-2-4-5.I. The higher accuracy is expected for Algorithm 2 since higher order term in Taylor Expansion is involved. The result is listed in Table 4.8 and is shown in Figure 4.14 together with data from Table 4.6.

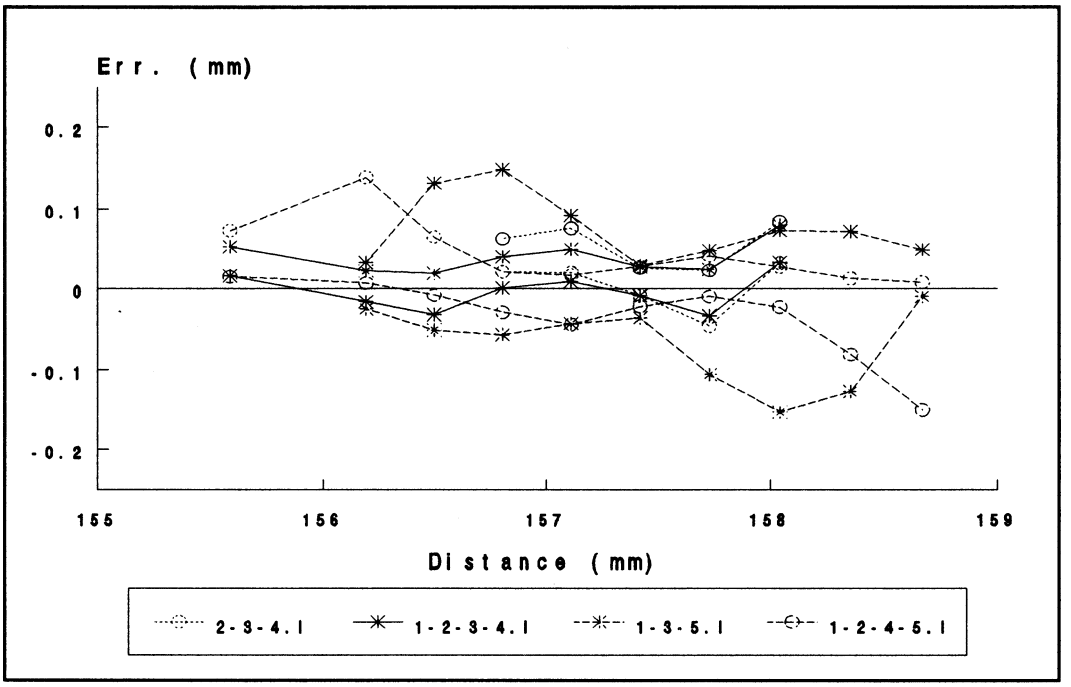


Figure 4.14 Algorithm 1 *vs* Algorithm 2.

It is observed that the 1-2-3-4.I scheme from Algorithm 2 performs best, with a comparatively large M-range and the lowest average error level over the range. The 1-2-4-5.I scheme has a slightly larger M-range but an apparently higher error level. One problem with Algorithm 2 is that, for "bad" points on the images (*e.g.* due to the glary surface points), the calculation under the square root in Eq.(3-29) is often negative. In practice, we assign to the bad point a large value that is beyond

the M-range, which is then removed through median filtering.

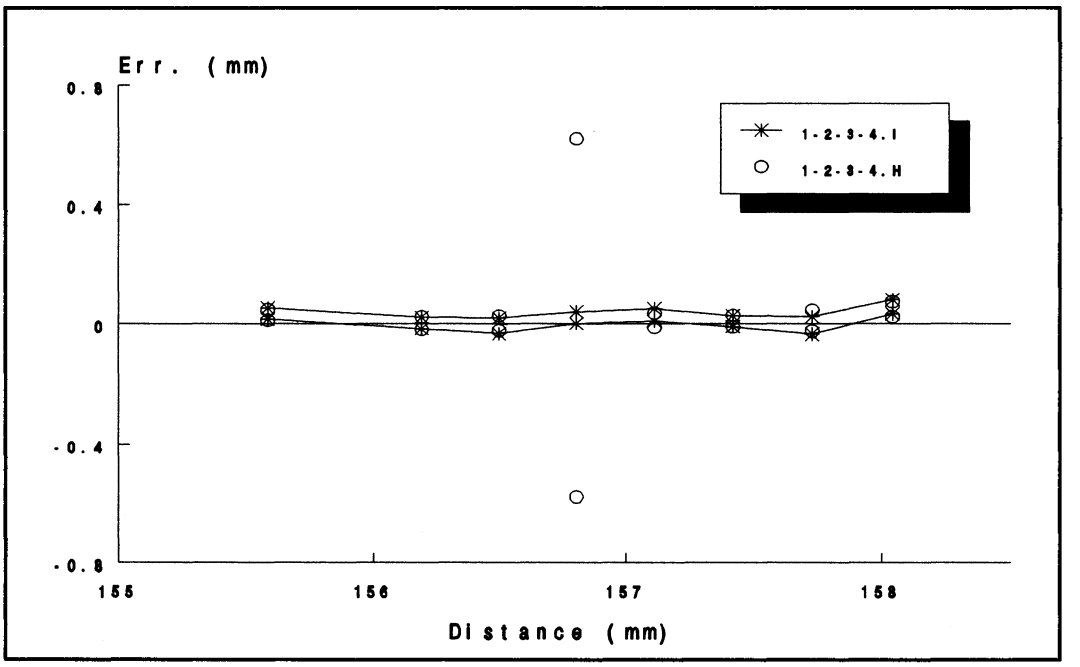
Table 4.8 Distance Estimates and Errors: Algorithm 2.

Target Posi. (mm)	1-2-3-4.I		1-2-4-5.I	
	Ave. Dist. (mm)	σ (mm)	Ave. Dist. (mm)	σ (mm)
154.981	—	—	—	—
155.584	155.619	**0.019**	155.628	**0.029**
156.191	156.194	**0.019**	156.264	**0.066**
156.497	156.490	**0.025**	156.526	**0.037**
156.804	156.824	**0.020**	156.800	**0.025**
157.112	157.142	**0.021**	157.099	**0.031**
157.421	157.430	**0.018**	157.424	**0.025**
157.731	157.726	**0.029**	157.746	**0.026**
158.043	158.100	**0.024**	158.045	**0.025**
158.356	—	—	158.322	**0.047**
158.670	—	—	158.599	**0.079**
159.302	—	—	—	—
159.939	—	—	—	—

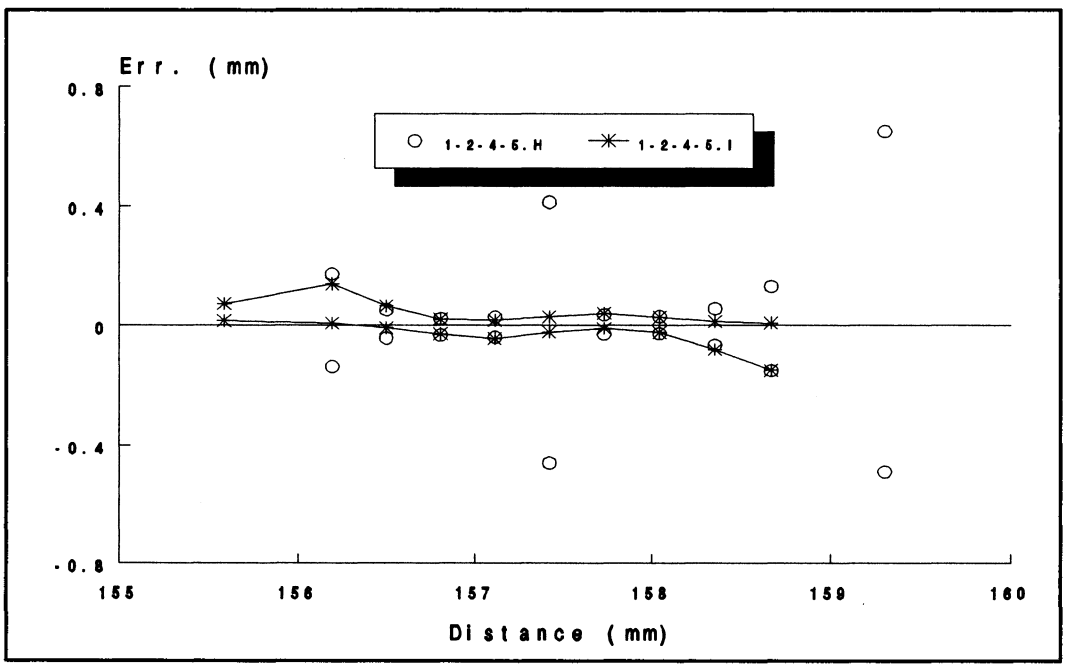
The 2-3-4.I scheme from Algorithm 1 is relatively simple and fast. For objects with small depth range, it is an alternative to Algorithm 2.

The hyperbola algorithm is also tested under two schemes. The resultant data are given in Table 4.9. Under either of the schemes, we observe a significant error at the centre of the I-range. On close examination, we attribute it to the inherent defect in the algorithm.

Algorithm 2 and the hyperbola algorithm are compared in Figure 4.15. With this and the result in Figure 4.14, we are able to claim the 1-2-3-4.I (or 2-3-4-5.I) to be the best amongst the present schemes available.



(a) 1-2-3-4.I *vs* 1-2-3-4.H.



(b) 1-2-4-5.I & 1-2-4-5.H.

Figure 4.15 Algorithm 2 *vs* Hyperbola Algorithm.

Table 4.9 Distance Estimates and Errors: Hyperbola Algorithm.

Target Posi. (mm)	1-2-3-4.H		1-2-4-5.H	
	Ave. Dist. (mm)	σ (mm)	Ave. Dist. (mm)	σ (mm)
154.981	154.607	2.875	155.385	6.921
155.584	155.614	**0.018**	155.832	3.992
156.191	156.194	**0.021**	156.208	**0.155**
156.497	156.499	**0.025**	156.502	**0.047**
156.804	156.823	**0.599**	156.800	**0.028**
157.112	157.124	**0.024**	157.107	**0.033**
157.421	157.430	**0.020**	157.397	**0.436**
157.731	157.743	**0.035**	157.734	**0.033**
158.043	158.092	**0.025**	158.044	**0.028**
158.356	158.535	0.346	158.351	**0.061**
158.670	158.806	7.765	158.660	**0.141**
159.302	—	—	159.381	**0.568**
159.939	—	—	—	—

4.4.2 Real Scene Tests

Three real scenes are tested and all surface points are processed (*i.e.* $L=1$). The objects are all placed near the central region of the calibrated range to ensure the minimal error level and the maximum range of measurement. One pixel size on these depth maps represents, in both orthogonal directions, about 0.0255mm in real world dimensions.

(1) Scene 1: U.S. Coin Surface

The maximum depth variation on the U.S. coin surface is measured to be

about 0.22mm. Two schemes (2-3-4.I and 1-2-3-4.I) are employed in estimating the surface depth and no difference in performance is observed between the two. In fact, the same performance is expected given such a depth range (refer to Figure 4.14). Two window sizes (10×10 & 16×16) are used and, from the previous testing results, the depth resolution ($\pm 3\sigma$) for the windows are estimated at about ±0.17mm and ±0.1mm respectively. The actual depth resolution could be lower than the estimated due to less spatial frequency content and excessive directional reflection from the metal surface.

Figure 4.16 (a) gives the depth map for 10×10 window and (b) for 16×16 window. In either case, the 3-D relief pattern on the coin surface is distinguished from the background. As seen, while the smaller window results in the noisier depth map and less depth resolution power, the larger window losses some pattern detail.

The corresponding intensity image is shown in (c). Several intensity images are involved in obtaining the depth map. By the *corresponding intensity image* we mean one of the intensity images that corresponds to $R = 1$ (refer to Section 1.1 and Eq.(3-2)). Median filtering with 5×5 window is applied onto the depth map to reduce the scattered noise.

(2) Scene 2: Spring Washers

A scene composed of two metal spring washers is tested under 1-2-3-4.I scheme and with 16×16 window. The result from this scene has been shown as an example in Figure 1.2. The cross-sections of the median filtered and unfiltered depth

maps have also been presented in Figure 4.5. As seen in Figures 1.2(b), there are some excessive glaring places on the edges of the washers.

Note that the metal surfaces in both scenes have not gone through any physical or chemical processing. It is believed that the results could be better should such processing be performed.

(3) Scene 3: An Inclined Plane

In Section 4.4.1, we obtain the relationship between the distance and the error in distance estimation at discrete target positions. For the 1-2-3-4.I scheme with 16×16 window, for example, σ and σ_e are estimated at about 0.039mm and 0.027mm respectively (refer to Eq.(4-10)). The error of measurement, δ, is then estimated from

$$\delta = \pm 3\sqrt{\sigma_e^2 + \sigma^2} \qquad (4-12)$$

at about ±0.14mm.

We now examine how the error varies continuously over the range of measurement. This time, the standard target is inclined slightly from its original vertical position. Since the depth variation of the inclined plane in the field of view is out of the M-range of 1-2-3-4.I scheme, two schemes, 1-2-3-4.I and 2-3-4-5.I, are used jointly to produce the depth map over such a depth range. The 16×16 window is applied and the resultant depth map (without median filtering) is shown in Figure 4.17(a). One of the intensity images is in (b), the focus condition on which, as seen,

varies constantly from the nearest surface point to the farthest in the field of view. Figure 4.18 provides a cross-section of the depth map. The standard deviation of the estimated distances from the real distances, measured on this depth map, is about 0.044mm (or $\delta \approx \pm 0.013$mm), which is very close to that obtained from the discrete estimation.

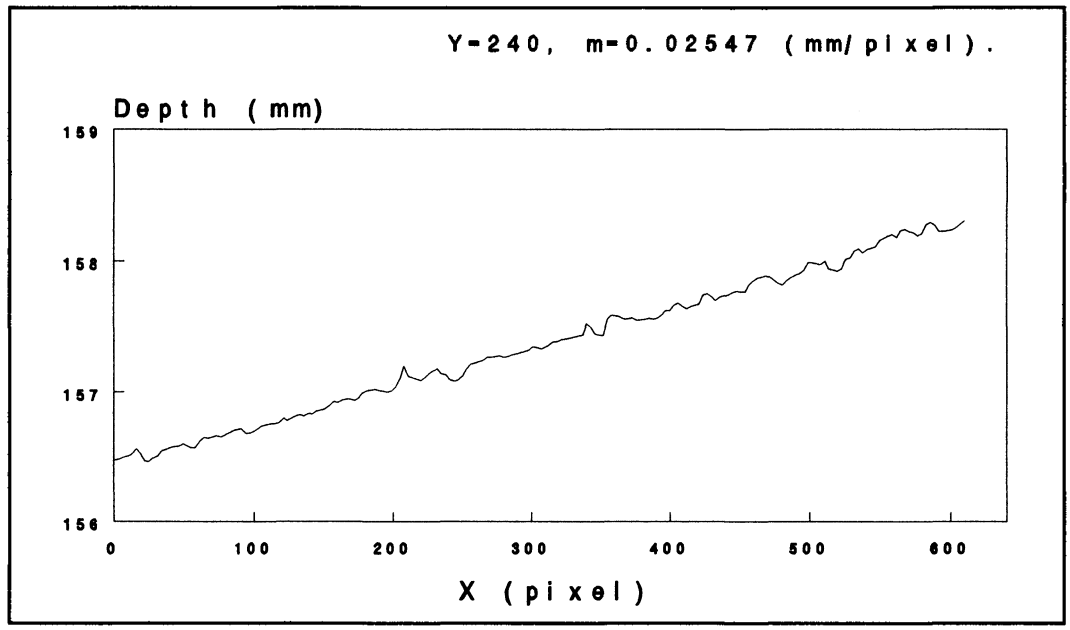


Figure 4.18 A Cross-Section of Depth Map of Scene 3.

It is necessary to note that no detectable field curvature is found in this experiment. We attribute the flat field to the high quality lens design and the relatively small field of view. The effects of image distortion and other error sources are briefed in Section 5.1.

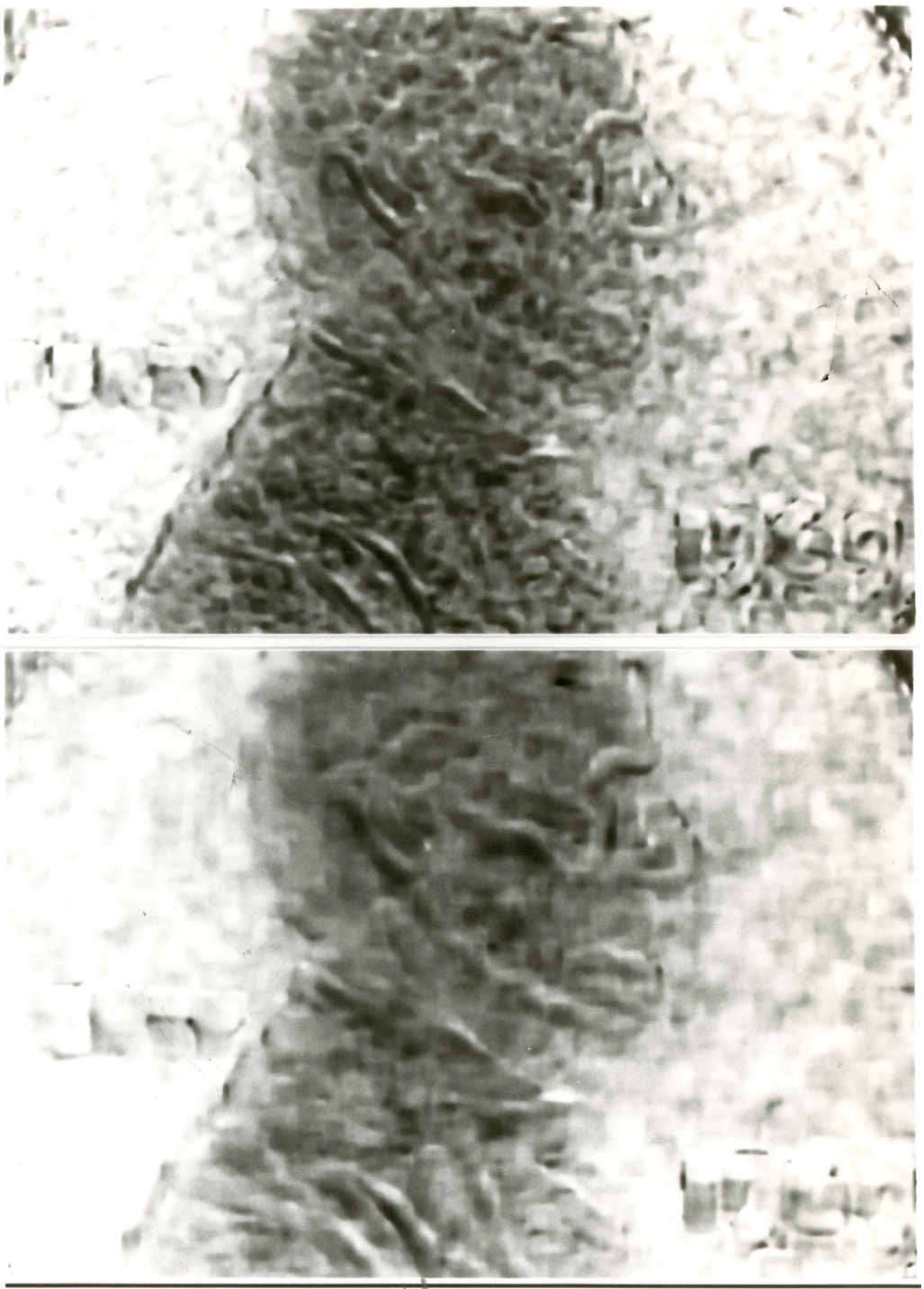


Figure 4.16 Scene 1: U.S. Coin Surface.
(a) Depth Map: 10×10 Window.
(b) Depth Map: 16×16 Window.

Figure 4.16 Scene 1: U.S. Coin Surface.
(c) Image of Brightness.

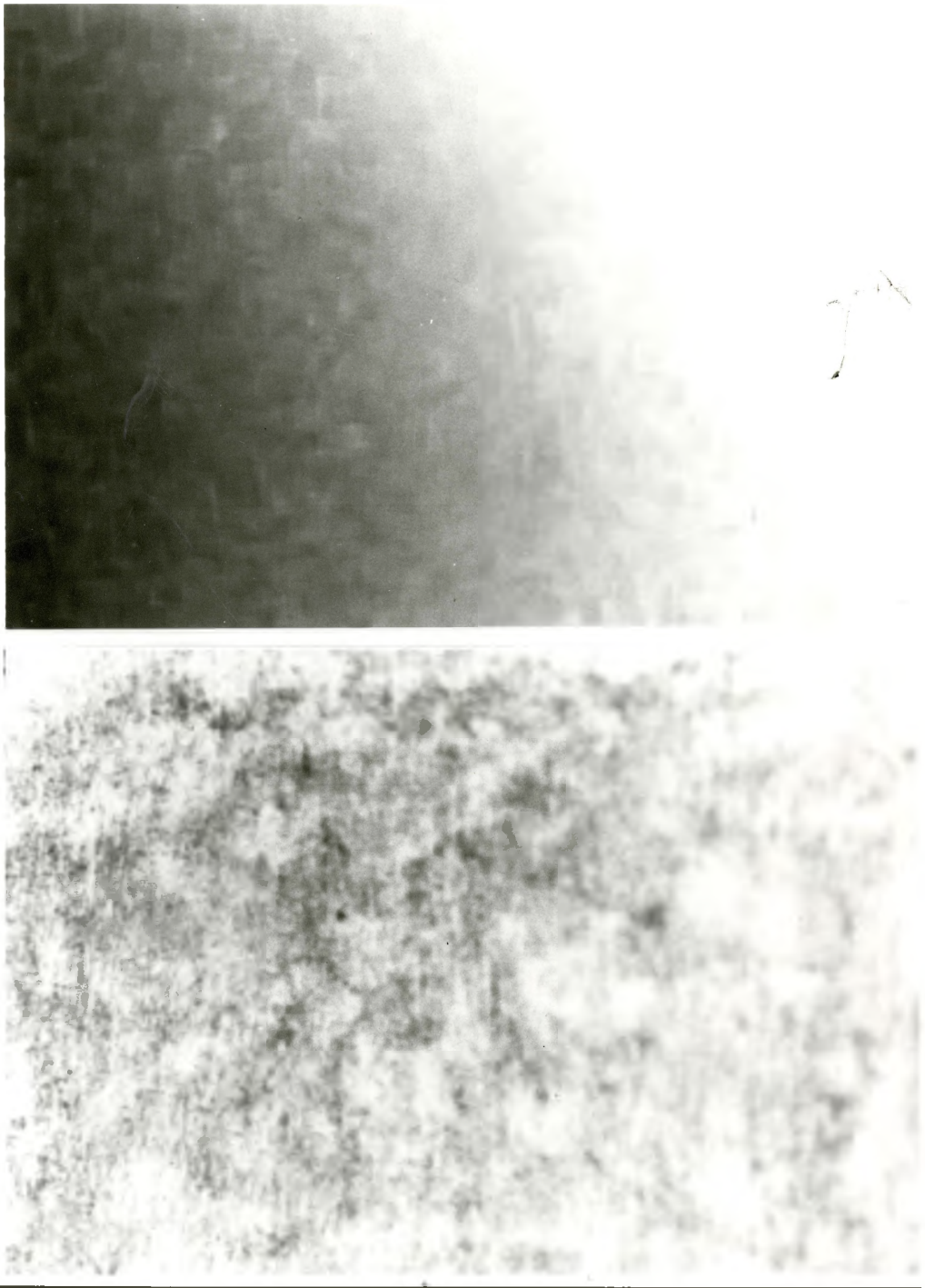


Figure 4.17 Scene 3: An inclined Plane.
(a) Depth Map.
(b) Image of Brightness.

CHAPTER 5

DISCUSSION

5.0 Introduction

In this chapter, we first discuss errors at different implementation levels. Then we analyze and compare the attained accuracies in depth acquisition from focus-based methods including this new technique. We finally discuss its advantage and limitation, and summarize the work.

5.1 Errors and Accuracies

Various error sources, some of which have been briefly discussed in the previous chapters, contribute to the error or the deformation of depth maps acquired from this technique. It is almost impossible to sort out individual error sources and carry out quantitative analysis. While it makes little sense just to specify these error sources without quantitative analysis, it may be helpful to categorize them according to where in the implementation procedure they appear to affect most:

Stage 1: Model Assumption

This technique is primarily based on the desirable attribute of $J(t)$. While its

validity has been theoretically justified, the model may not be strictly valid in a real system, due to the presence of factors like *vignetting* and *transversal optical aberrations*. These factors change more or less with distance and affect the assumed symmetry of focal gradient function. In most cases, the changes are minor and often negligible.

Stage 2: Model Approximation

Without question, to approximate the focal gradient function, in one form or another, introduces *theoretical error*. As seen by the experimental results, the 4th order Taylor Expansion provides a satisfactory approximation for the model.

Stage 3: Calibration

Error of distance measurement causes deviations of the estimated distances. The error is mainly due to the limited *resolution of measuring tools* and human *error in reading* amidst calibration, and can be statistically reduced through multiple measurements. The effect of this error changes with system magnifications. For example, the 2μm resolution of the micrometer fixed to the CCD camera in this implementation is amplified to about 7μm in object-space.

Stage 4: Implementation

Window operation is required to calculate the values of the focal gradient function, which are then used for distance estimation. *Thermal noise* and *image quantization* and *discretization* have a direct bearing on producing random errors in window calculation. The sharpness of the focal gradient extreme depends primarily upon the richness of *spatial frequency content* in the window, and the flat-peaked focal

gradient is more vulnerable to the random noise. Window position error (mainly due to the *error in locating the optical axis*) causes non-correspondence amongst the windows. It is believed that the lack of spatial frequency content on the object surface plays a key role in determining the severity of the error in distance estimation at this stage.

Stage 5: Range Image Mapping

The *field curvature* results in the axial deformation of the depth map. Note that the *image distortion* does not add extra error to distance estimates in the original depth map of Z-$\theta\phi$ form, for the images are distorted in the same way and to almost the same degree, and thus no matching problem exists amongst them. However, the mapping to Cartesian system as described in Eq.(1-1), making use of the (distorted) 2-D coordinates on the depth map, introduces distortion in depth. Given necessary optical data, the distorted depth map can be corrected. In this experiment, we discern within the field of view neither of the effects that is beyond the measurement errors and noise.

Discussions on the accuracy of depth estimation from focus can be found in much of the early work. Pentland (1987) and Das (1989) compared the theoretical accuracies from focal accommodation and stereopsis and concluded they are comparable with each other over a certain range of distance. Krotkov (1987) and Das (1989) provided formulae for calculating the *depth of field* from geometrical optics. Generally speaking, the depth of field is a range of distance in object-space,

within which the object points are indiscernibly imaged onto the image receptor.

The *depth of focus*, $\Delta l'$, is a range of distance about the focal point in image-space, determined by Rayleigh quarter-wave limit (Kingslake, 1983). It is widely used as a measure of tolerance for aberration in optical systems. For a general system, we have

$$\Delta l' = \pm \frac{\lambda}{8 n' \sin^2 \frac{U'}{2}} \quad or \quad \frac{\lambda}{4 n' \sin^2 \frac{U'}{2}} \qquad (5\text{-}1)$$

where λ is the wave length; n' is the refractive index in image-space; U' denotes the half aperture angle in image-space. For small aperture, we have

$$\sin U' \approx \frac{D}{2 l'} \qquad (5\text{-}2)$$

where D is the diameter of the aperture and l' the image distance.

The corresponding range to the depth of focus in object-space is, assuming the system is in air:

$$\Delta l_f = \pm 2\lambda \cdot \frac{l^2}{D^2} \qquad (5\text{-}3)$$

where l is the object distance.

This equation reflects how the depth resolution, or the depth of field, changes with distance and aperture in a diffraction-limited system. Although the cause for the depth error in real systems is more complicated, the basic relationship in Eq.(5-3) should remain. Alternatively, Δl_f can be as a measure of evaluating accuracies in

depth estimation, attained from different techniques and system configurations.

For instance, in this implementation, the average distance $l=157.42$mm, $D=19.6$mm, and $\Delta l_f=\pm 0.063$mm (assume $\lambda=0.5\mu$m). The average error of measurement for 16×16 window is ±0.13mm, which is about *twice* as much as the Δl_f. For large windows (*e.g.*, ≥80×80), $\sigma\approx\sigma_{min}=0.01$mm, and the minimal error is estimated at about ±0.086mm from Eq.(4-12) (assuming the same σ_e). Note that the depth resolution (random depth error) with the large window size is $\pm 3\sigma_{min}$, only about *half* of the Δl_f.

Table 5.1 Accuracies from Focus-Based Methods.

Researcher(s)	Accuracy Reported	l	D	Δl_f
Rioux, *et al.* (1986)	resolution: 1mm	1m	30mm	±1mm
Grossman (1987)	$\sigma=1.25$cm	1m		
Krotkov (1987)	$\sigma=0.6$~1.6%	1.5m~3m	58mm	±0.67~2.67mm
Engelhardt (1988)	mean error: ±1.5m	500mm	20mm	±0.63mm
Pentland, *et al.* (1989)	$\sigma=2.5$~6%			
Cardillo, *et al.* (1991)	average error: 4.5mm	750mm	20mm	±1.4mm
Lai, *et al.* (1992)	error<5%	<163cm		
Nayar (1992)	mean error: 7.86μm			
— (1992)	$\sigma_{min}=0.01$mm, $\delta=\pm 0.13$mm	157mm	20mm	±0.06mm

Table 5.1 lists some reported accuracies from focus-based techniques. The concerned system parameters, if available, and the corresponding Δl_f are also listed in the table for reference. The wave length $\lambda=0.5\mu$m is assumed in calculating Δl_f. Different terms were used to describe the attained accuracies. Unless it is unmistakably recognized as the standard deviation, which is represented by σ in the

table, the accuracy is presented the way as it was in the literature.

It is observed that only the accuracies from two active methods are comparable to that from this technique, with the "resolution" from Rioux (1986) being about the same as the Δl_f and the "mean error" from Engelhardt (1988) about two times the Δl_f. However, the surface albedo cannot be recovered and only distances from limited surface points are extracted with these methods, due to the structured illumination patterns. While the overall superiority of our technique is not claimed for lack of tests over a large range of distance, it is believed that the technique is among the best in approaching the physical limit under given system configuration.

5.2 Conclusion

This technique is fully passive and parallel: both natural scene radiance and surface depth are recovered simultaneously without physical scanning. Modelling the gradient of focus instead of the PSF avoids, in principle, errors due to window border effect and other practically constant error sources. The depth acquisition becomes a procedure of parameter estimation for the focal gradient function. Within a certain range, the approximation of 4th degree for the model is adequate and the distances can be extracted with accuracy.

The major disadvantage of this technique is its limited range of measurement. While the problem can be solved by joining together several estimations over a larger range (as we do for Scene 3 in Section 4.4.2), the execution time increases. Secondly,

unlike other SfD methods that only need one or two images, this technique uses up to four images to obtain a constrained depth estimation. Consequently, more processing time is required. Although fast algorithm and optimized programs can be used, it seems that for real-time application, special parallel processing hardware and fast motion control system would have to be employed.

Other related work worthy to be considered in the future include:

1) Test objects over a large range of distance, which may involve calibrating a large number of object/image positions.

2) Employ focus sharpness criteria (FSC) other than the energy measure, especially in finding a more accurate analytical model for the focal gradient function.

3) Since the field of view is limited for the commercial Nikon lens, this technique would certainly benefit from a specially designed and fabricated system that has large field of view and comparatively small aberrations.

4) Develop a fully automated focus-based ranging system for 3-D object surface reconstruction in computer vision and industrial applications.

In summary, a new image-based technique for depth extraction from focus is proposed. The technique is implemented on a simple opto-digital image processing system. Accurate distance estimates are obtained that are comparable to the results from existing active focus-based methods.

REFERENCES

Besl, P.J. "Range Image Sensors", *Advances in Machine Vision: Application & Architecture*, New York: Springer-Verlag, 1988, pp.1-117.

Born, M. and Wolf, E. *Principle of Optics*, London: Pergamon, 1965.

Cardillo, J. and Sid-Ahmed, M.A. "3-D Position Sensing Using a Passive Monocular Vision System", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol.13, Aug. 1991, pp.809-813.

Darrell, T. and Wohn, K. "Pyramid Based Depth from Focus", *Proc. IEEE CVPR'88*, Ann Arbor, MI, June 1988, pp.504-509.

Das, S. and Ahuja, N. "Integrating Multiresolution Image Acquisition and Coarse-to-Fine Surface Reconstruction", *Proc. of Workshop on Interpretation of 3D Scenes, IEEE*, November 1989, pp.9-15.

Engelhardt, K. "Acquisition of 3-D Data by Focus Sensing", *Applied Optics*, Vol.27, November 1988, pp.4684-4489.

Goodman, J.W. *Introduction to Fourier Optics*, New York: McGraw-Hill, 1968.

Grossman, P. "Depth from Focus", *Patt. Recogn. Lett.*, Vol.5, January 1987, pp.63-69.

Horn, B.K.P. "Focusing", MIT Project Mac, AI Memo No.160, May 1968.

Horn, B.K.P. *Robot Vision*, Cambridge: MIT Press, 1986.

Huang, T.S. "A Fast Two-Dimensional Median Filtering Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing*, Vol.27, February 1979, pp.13-18.

Hwang, T. *et al.* "A Depth Recovery Algorithm Using Defocus Information", *Proc. IEEE CVPR'89*, San Diego, CA, June 1989, pp.476-481.

Jarvis, R.A. "Focus Optimization Criteria for Computer Image Processing", *Microscope*, Vol.24, 2nd Quarter 1976, pp.163-180.

Jarvis, R.A. "A Perspective on Range Finding Techniques for Computer Vision", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol.5, March 1983, pp.122-139.

Kanade, T. *Three-Dimensional Machine Vision*, Netherlands: Kluwer Academic, 1987.

Kingslake, R. *Optical System Design*, New York: Academic, 1983.

Krotkov, E. "Focusing", *Int. J. Comput. Vis.*, Vol.1, November 1987, pp.223-237.

Lai, S. *et al.* "A Generalized Depth Estimation Algorithm with a Single Image", *IEEE Trans. Patt. Anal. Machine Intell.*, Vol.14, April 1992, pp.405-411.

Nayar, S.K. "Shape from Focus System for Rough Surfaces", *Proc. Image Understanding Workshop*, Darpa, January 1992, pp.593-606.

Pentland, A.P. "A New Sense for Depth of Field", *IEEE Trans. Patt. Anal.*

Machine Intell., Vol.9, July 1987, pp.523-531.

Pentland, A.P. "Progress toward A Simple, Parallel Vision Machine", *Optics News*, Vol.15, May 1989, pp.9 & 26-32.

Pentland, A.P. *et al.* "A Simple, Real-Time Range Camera", *Proc. IEEE CVPR'89*, San Diego, CA, June 1989, pp.256-261.

Rioux, M. and Blais, F. "Compact Three-Dimensional Camera for Robotic Applications", *J. Opt. Soc. Am. A*, Vol.3, September 1986, pp.1518-1521.

Schlag, J.F. *et al.* "Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control", Carnegie-Mellon Univ., CMU-RI-TR-83-14, August 1983.

Shirai, Y. *Three-Dimensional Computer Vision*, New York: Springer-Verlag, 1987.

Subbarao, M. "Direct Recovery of Depth-Map", *Proc. IEEE Comput. Soc. Workshop Comput. Vision*, Miami Beach, FL, December 1987, pp.58-65.

Subbarao, M. and Natarajan, G. "Depth Recovery from Blurred Edges", *Proc. IEEE CVPR'88*, Ann Arbor, MI, June 1988, pp.498-503.

Subbarao, M. "Parallel Depth Recovery by Changing Camera Parameters", *Proc. IEEE 2nd Int. Conf. Computer Vision*, Tampa, FL, December 1988, pp.149-155.

Subbarao, M. "Efficient Depth Recovery through Inverse Optics", *Machine Vision for Inspection And Measurement*, New York: Academic Press, 1989, pp.101-126.