Practical Temporal Psychovisual Modulation with Liquid Crystal Devices

PRACTICAL TEMPORAL PSYCHOVISUAL MODULATION WITH LIQUID CRYSTAL DEVICES

BY

FANGZHOU LUO, B.E.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

Master of Applied Science (2017) McMaster University
(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE: Practical Temporal Psychovisual Modulation with Liquid Crystal Devices

AUTHOR: Fangzhou Luo
B.E., (Department of Electronic Engineering and Information Science)
University of Science and Technology of China, Hefei, China

SUPERVISOR: Dr. Xiaolin Wu

NUMBER OF PAGES: xii, 38

Dedicated to my parents

Abstract

Temporal Psychovisual Modulation (TPVM) is a new paradigm of computational display, which can concurrently exhibit many different views on a common display. These views are decomposed into a set of atom frames, and the atom frames are modulated by users' liquid crystal (LC) light modulation glasses. Due to the limited refresh rate of LC glasses, a practical TPVM multiview display system can only support two to three atom frames. Such a small number of atom frames are not sufficient to a multiview application with a high image quality. Therefore, the main technical challenge before TPVM is how to support as many viewers as possible while maintaining an acceptable perceptual quality, using only a small number of atom frames. In this thesis, we develop two approaches to meet the challenge.

The first approach is to exploit the sparsity of multiview images to be displayed. One example is the sequence of depth of field (DOF) images. Those images not only provide a strong depth cue, but also have a sparsity structure. That structure allows the DOF images to be reconstructed from a small number of atom frames. We prove that property theoretically. Experimental results also agree well with the proof.

The second approach is to exploit the well-known property of rapidly decreasing visual acuity from fovea to peripheral vision. The strategy is to exhibit different concurrent views at highest quality in viewers' focused regions, while allowing graceful image quality degradation in regions of peripheral vision. This is achieved by a novel fovea weighting algorithm that optimizes for subjective quality. We find the proposed algorithm improves viewers' perceptual quality significantly, especially when the TPVM multiview display system only has a small number of atom frames.

Acknowledgements

Foremost, I would like to thank my supervisor, Professor Xiaolin Wu, for his patient guidance, encouragement and advice during my research. This thesis would not have been possible without his support.

I would also like to thank my labmates for your discussion, cooperation and friendship.

Finally, I would like to thank my parents for supporting me throughout my life.

Notation and abbreviations

AR	augmented reality
CRT	cathode ray tube
DOF	depth of field
HMD	head-mounted display
HVS	human visual system
LC	liquid crystal
LCD	Liquid-crystal display
LED	Light-emitting diode
MR	mixed reality
NMF	non-negative matrix factorization
OLED	organic light-emitting diode
POCS	projection to convex set
RMSE	root mean square error

ROI	region of interests
SVD	singular value decomposition
TPVM	temporal psychovisual modulation
VR	virtual reality

Contents

Abstract **iv**

Acknowledgements **vi**

Notation and abbreviations **vii**

1 Introduction **1**

2 TPVM and other computational displays **4**

2.1 TPVM 5

2.2 Virtual reality head-mounted display 7

2.3 Augmented/mixed reality head-mounted display 8

2.4 Tensor Displays 10

2.5 Magic Leap's head-mounted display 11

3 Depth of Field Image Sequences **13**

3.1 The motivation and definition of DOF image sequences 14

3.2 Sparsity of DOF image sequences 15

3.3 Experimental results 21

4 Fovea Weighting of TPVM 26

4.1 Fovea weighting 27

4.2 Experimental results 29

5 Conclusions 33

List of Figures

2.1 TPVM multiview display system 6

2.2 Virtual reality head-mounted display 7

2.3 Augmented/mixed reality head-mounted display 9

2.4 Image formation mechanism of tensor display 10

2.5 Fiber Scanning Display 12

3.1 An example of DOF image sequence. 14

3.2 The maximum root mean square error of row vectors of $G - \tilde{G}$ if focal distance d_n changes from $10C_F$ to $1000C_F$ in a DOF image sequence of $N = 1000$ images. 20

3.3 TPVM outputs of DOF sequence in classroom scene with far focus. . 23

3.4 TPVM outputs of DOF sequence in classroom scene with near focus. 23

3.5 TPVM outputs of DOF sequence in the pavilion scene with far focus. 24

3.6 TPVM outputs of DOF sequence in the pavilion scene with near focus. 24

3.7 TPVM outputs of DOF sequence captured by Lytro Illum camera. . . 25

4.1 Results of fovea weighting of TPVM in classroom scene. The red circles indicate the centers of regions of interest of different users. The PSNRs reported are fovea weighted PSNR. 31

4.2 Results of fovea weighting of TPVM in the pavilion scene. The red circles indicate the centers of regions of interest of different users. The PSNRs reported are fovea weighted PSNR. 32

Chapter 1

Introduction

Recent years have seen intensified research on and burgeoning commercial interests in a new generation of computational displays, driven by a wide range of virtual, augmented and mixed reality applications in diversified fields from man-machine interactions, medicine, entertainment, to automobile, etc (Masia *et al.*, 2013). Temporal Psychovisual Modulation (TPVM) (Wu and Zhai, 2013) is a new paradigm of computational display, which can concurrently exhibit many different views on a common display surface. These views are decomposed into a set of atom frames to be displayed at a high frame rate exceeding the critical flicker frequency of 60 Hz for human eyes. Different users perceive their own views through liquid crystal (LC) light modulation glasses. The LC glasses, synchronized with the high-speed display, can regulate how much of the light energy of each atom frame to pass through and reach retina, namely, perform amplitude modulation of atom frames, so that the human visual system (HVS) can fuse these modulated atom frames into desired images.

Although the number of atom frames can be any positive integer theoretically, its feasible range is severely limited by the relatively low speed of active LC light modulation glasses. The speed of the off-the-shelf LC glasses of grey levels is difficult to exceed 180Hz. In other words, in the current state of the art, a practical TPVM multiview display system can only support two to three atom frames. Such a small number of atom frames are not sufficient to a multiview application with a high image quality. Therefore, the main technical challenge before TPVM is how to support as many viewers as possible while maintaining an acceptable perceptual quality, using only a small number of atom frames. In this thesis, we develop two approaches to meet the challenge.

The first approach is to exploit the sparsity of multiview images to be displayed. One example is the sequence of depth of field (DOF) images, which pertaining to continuously varying focal distance but with the position, angle and aperture of the camera fixed. Those images not only provide a strong depth cue, but also have a sparsity structure. That structure allows the DOF images to be reconstructed from a small number of atom frames. We prove that property theoretically. Experimental results also agree well with the proof.

The second approach is to exploit the well-known property of rapidly decreasing visual acuity from fovea to peripheral vision. We propose a spatially weighted optimization algorithm for TPVM based on viewers' real time region of interests (ROI) information. The strategy is to exhibit different concurrent views at highest quality in viewers' focused regions, while allowing graceful image quality degradation in regions of peripheral vision. This is achieved by a novel fovea weighting algorithm that optimizes for subjective quality. We find the proposed algorithm improves viewers' perceptual quality significantly, especially when the TPVM multiview display system only has a small number of atom frames.

The remainder of this thesis is structured as follows. Chapter 2 outlines the principle and architecture of TPVM and other state-of-the-art computational displays, points out advantages of TPVM, and stresses the necessity of continuing development of it. Chapter 3 presents our findings on the DOF image sequences and their sparse property. Chapter 4 presents the fovea weighting algorithm for optimizing multiuser perceptual quality, and Chapter 5 concludes.

Chapter 2

TPVM and other computational displays

The concept of display has a history that predates the modern flat-panel display by several millennia. People have a strong need to present information, even if they can only write in a stone. With the rapid development of display technologies, they become more automated and further clearer. The mechanical display, like Split-flap display and Flip-disc display are developed first. Then electronic display, like Cathode ray tube (CRT), Light-emitting diode (LED) and Liquid-crystal display (LCD) are developed not long after, and the modern flat-panel display is evolved from these primitive electronic display.

Nowadays, the flat-panel display is stable and mature enough, but people also have higher demands for their user experience. For example, people want more immersive visual experience, more accessible information, more interaction with the physical world, and more collaboration with other people. Satisfying these needs is an important task of modern computational displays. Computational display technology is one of the hottest topics in the graphics community today (Masia *et al.*, 2013). They can provide a magical visual experience by adding computational power to optics. In this chapter, we outline the principle and architecture of TPVM and other state-of-the-art computational displays, points out advantages of TPVM, and stresses the necessity of continuing development of it.

2.1 TPVM

In the TPVM multiview display system (Wu and Zhai, 2013; Zhai and Wu, 2014), as depicted in Figure 2.1, K concurrent views are decomposed by non-negative matrix factorization into a set of atom frames, i.e., basis images. These atom frames are displayed at a high frame rate exceeding the flicker frequency. K viewers watch the screen through liquid crystal (LC) light modulation glasses that are synchronized with the display and perform temporal amplitude modulation of the atom frames to generate desired images for different viewers. The TPVM display system simplifies the VR end user device from a head-mounted display to light, simple LC glasses. As the LC glasses are see through, the VR participants can conduct face-to-face communications or even body-to-body interactions. The co-presence of multiple users in virtual environment is achieved via perceptual fusion of the perspective-correct virtual world and the participants' own physical proximity.

At the heart of the multiuser display system is a problem of non-negative matrix factorization (NMF) (Lee and Seung, 1999). Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_K)$ be the K target images to be concurrently displayed to different viewers. The $S \times K$ matrix $\boldsymbol{Y}$, where S is the number of pixels in each target image, needs to be decomposed

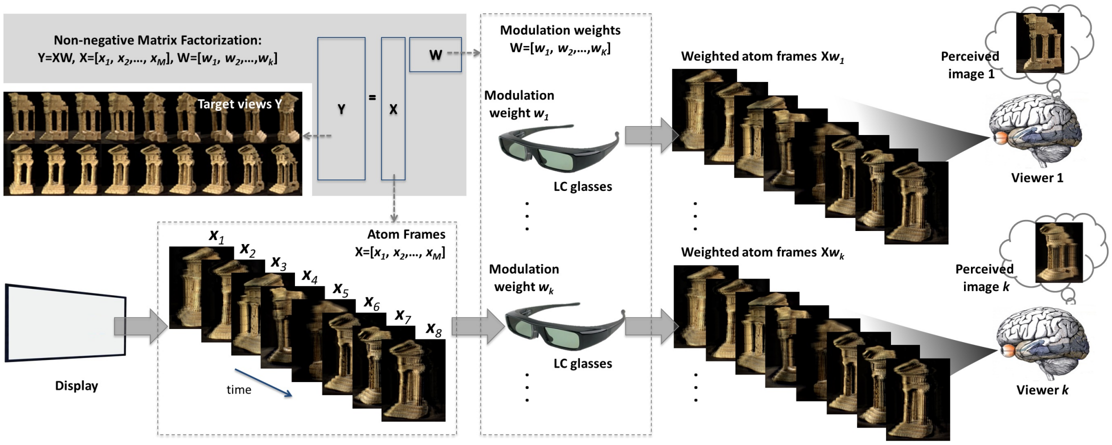


Figure 2.1: TPVM multiview display system

into $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}$, with the $S \times M$ matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M)$ being the set of atom frames and the $M \times K$ matrix $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_K)$ being the K modulation coefficient vectors corresponding to the K target images. The resulting atom frames $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ are displayed at high frame rate and then temporally modulated by active LC glasses according to weights $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_K$. This optoelectronic display-glass coupling and the temporal fusion mechanism of HVS jointly render the K concurrent target images $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_K$ as different linear combinations of the $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ atom frames. By now one can appreciate that in the TPVM paradigm, all heavy computations involved in multiview generation are delegated to a central server. End user devices become inexpensive, light LC glasses that are controlled by a modulation vector that only consumes a negligible bandwidth.

In practice, the image decomposition underlying TPVM has to respect a condition of non-negativity, because the light energy emitted by the display cannot be negative, and active LC glasses can only implement modulation weights between 0 and 1. Therefore, the introduced display system needs to solve the following problem of

NMF (Lee and Seung, 2001)

$$\min_{\boldsymbol{X},\boldsymbol{W}} \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W}\|_F^2, \quad \text{subject to} \quad 0 \leq \boldsymbol{X}, \boldsymbol{W} \leq 1 \tag{2.1}$$

with $\boldsymbol{Y} \in \Re^{S \times K}$, $\boldsymbol{X} \in \Re^{S \times M}, \boldsymbol{W} \in \Re^{M \times K}$, and where $\| \cdot \|_F$ is the Frobenius norm; $\leq$ operates on each element of the matrices.

2.2 Virtual reality head-mounted display

Modern virtual reality head-mounted displays (HMD), like Oculus Rift (Oculus, 2016), PlayStation VR (Sony, 2016), HTC Vive (HTC, 2016) and Samsung Gear VR (Samsung, 2015), have very similar structures. As sketched in Fig 2.2, they all have a stereoscopic head-mounted display which provides separate images for each eye, head motion tracking sensors and maybe eye tracking sensors.

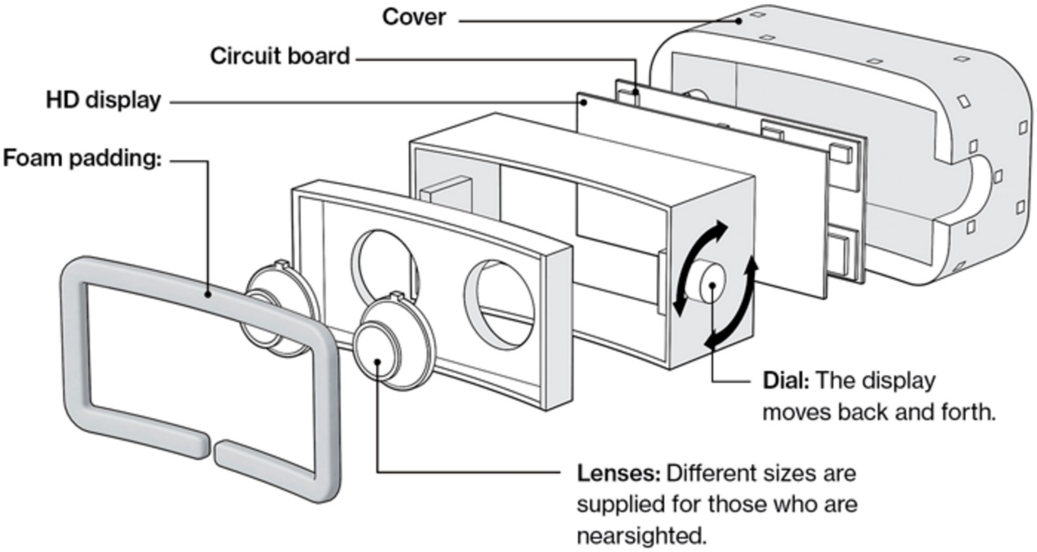

Cover
Circuit board
HD display
Foam padding:
Dial: The display moves back and forth.
Lenses: Different sizes are supplied for those who are nearsighted.

Figure 2.2: Virtual reality head-mounted display

Their technical specifications are also similar. Taking Oculus Rift as an example, it uses an organic light-emitting diode (OLED) panel for each eye, each having a resolution of $1080 * 1200$, a refresh rate of 90 Hz rate, and 110° field of view. It also has full 6 degree of freedom rotational and positional tracking which is precise, low-latency, and sub-millimeter accurate.

A head-mounted display can easily create an immersive environment for a user, which is similar to the real world. However, it also isolates him/her from others. HMD cannot give one the perception of corporal co-presence with others in the same virtual environment. In many applications of multiuser collaborative VR, such as surgical planning, training for manned space exploration, exercising for emergency in hostile environment, etc., participants need to act together in real physical proximity rather than connected in cyber space. TPVM multiview display system, by contrast, is ideally suited for multiuser collaborative VR.

2.3 Augmented/mixed reality head-mounted display

Augmented reality (AR) and mixed reality (MR) head-mounted display, sometimes known as ”Smartglasses”, are developing rapidly. AR is a view of a physical, real-world environment whose elements are augmented by external data, while MR is the merging of real and virtual worlds to produce new environments where physical and digital objects co-exist and interact in real time. Google Glass (Google, 2014) is a typical instance of AR head-mounted display, and Microsoft HoloLens (Microsoft, 2016) is a typical instance of MR head-mounted display. The goals of the two Smartglasses

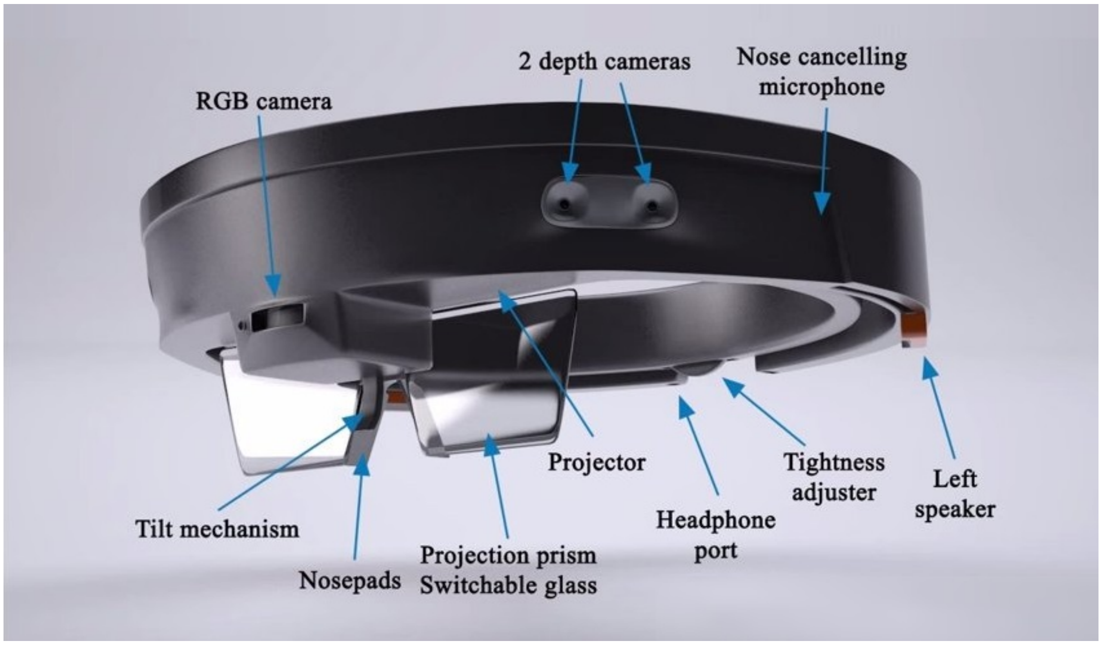


Figure 2.3: Augmented/mixed reality head-mounted display

are different, but they have similar structures as display systems. They all have the capability of reflecting projected images to the user's eyes while allowing them to see through the glasses. As an example, the structure of HoloLens is shown in Fig 2.3.

Although augmented/mixed reality head-mounted display seems like a good scheme of multiuser collaborative VR, they also have their drawbacks. They have low resolutions (360p for Google Glass and 720p for HoloLens), narrow fields of view (14° for Google Glass and 30° for HoloLens) and high prices ($1500 for Google Glass and $3000 for HoloLens). Their drawbacks limit their commercial use in multiuser collaborative VR.

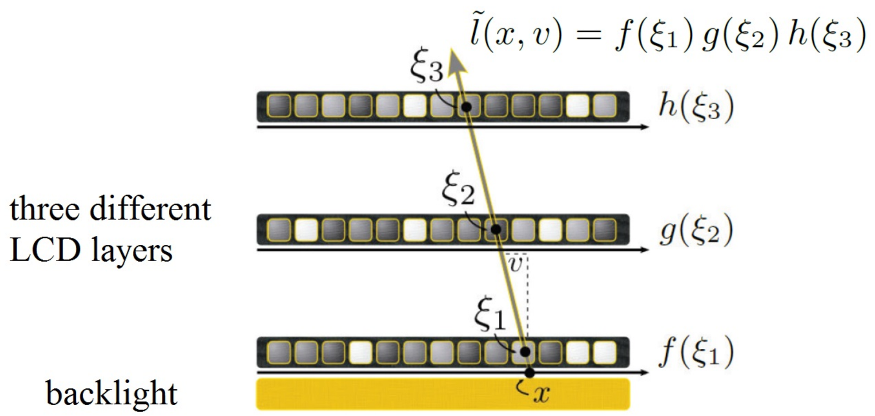

$\tilde{l}(x,v) = f(\xi_1)\,g(\xi_2)\,h(\xi_3)$
ξ_3
$h(\xi_3)$
three different LCD layers
ξ_2
$g(\xi_2)$
v
ξ_1
$f(\xi_1)$
backlight
x

Figure 2.4: Image formation mechanism of tensor display

2.4 Tensor Displays

Tensor displays (Wetzstein *et al.*, 2012) are a family of light field displays represented by MIT Media Lab, with the potential of resolving the vergence-accommodation conflict (Kramida, 2016). They share a degree of similarity with TPVM in optical image formation mechanism and in computational aspect: both use LC glasses as spatial light modulator to generate many concurrent views, and both decompose images into a set of atom frames. Unlike TPVM that uses only one LC modulation layer, the tensor display uses two or more LC modulation layers to gain multiview capability. As sketched in Fig 2.4, the tensor display approximates the light field by solving the following optimization problem:

$$\min_{\boldsymbol{F},\boldsymbol{G},\boldsymbol{H}} \|\boldsymbol{L} - \boldsymbol{W} \otimes (\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{f}_m \circ \boldsymbol{g}_m \circ \boldsymbol{h}_m)\|, \quad \text{subject to} \quad 0 \leq \boldsymbol{f}_m, \boldsymbol{g}_m, \boldsymbol{h}_m \leq 1 \tag{2.2}$$

where $\boldsymbol{L}$ is the light field tensor, $\boldsymbol{W}$ is the physical restriction tensor, $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ are transmittance vectors of three different LCD layers. In TPVM each user has his/her own LC classes that perform light amplitude modulation independent of other users. But the LC modulation layers of the tensor display are responsible for approximating all views of the light field. This distinction in design leads to different computation models: nonnegative matrix factorization for TPVM, as in Eq 2.1, but nonnegative tensor factorization for the tensor display, as in Eq 2.2.

The tensor display can generate glasses-free 3D views, which is its main advantage over TPVM. But because all concurrent views generated by the tensor display are physically formed by layers of spatial light modulators, or mathematically the results of a tensor decomposition, they are severely constrained. In general, the tensor display cannot reproduce the 4D light field of an arbitrary 3D scene. In implementation, even small errors in layer registration can cause severe crosstalk, destroying image clarity and narrowing the effective field of view. Also, the use of multiple LC layers deprives light efficiency, generating dim and low contrast images. In contrast, in the TPVM display system, as each user uses an independent, single LC layer, layer registration and crosstalk become nonissues, images should be brighter, sharper and cleaner than those of tensor displays.

2.5 Magic Leap's head-mounted display

A mysterious head-mounted display is being developed at Magic Leap, a US startup company. They posted a series of videos of their product online in 2016. In one of the most famous video, a whale appeared to breach the floor, leap high in the air and come crashing down with water flying everywhere. These videos created a sensation in the

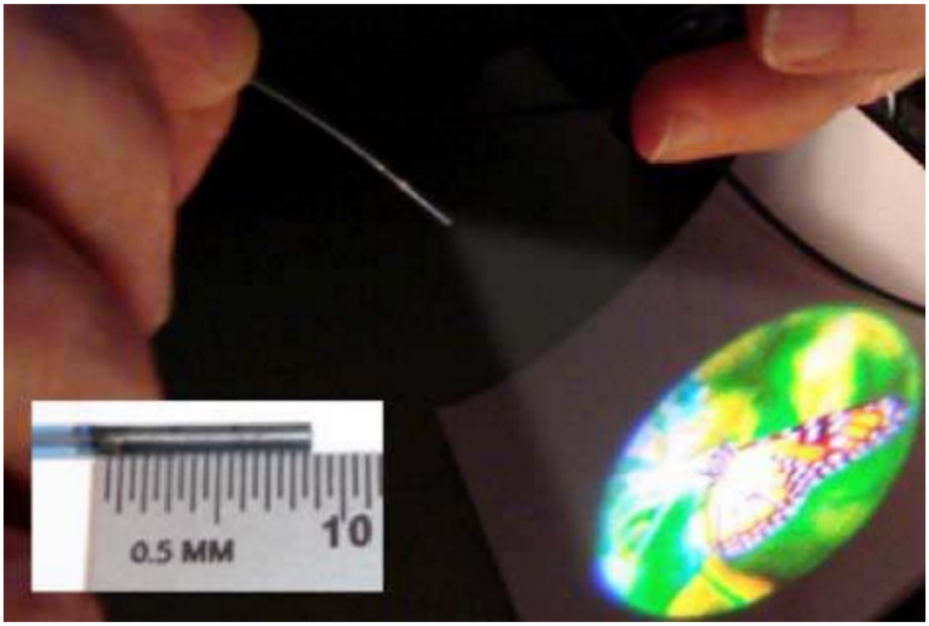


Figure 2.5: Fiber Scanning Display

world, and led people to believe that Magic Leap already made a major breakthrough in display technology. However, it turns out that all those videos are made by other special effects companys, rather than shot directly by camera. Magic Leap has not demonstrated a prototype till today.

There is widespread speculation that Fiber Scanning Display (Crossman-Bosworth *et al.*, 2006; Schowengerdt *et al.*, 2010) is the key in their system, which could shine a laser through a fiber optic cable that moves rapidly back and forth to draw images out of light, as shown in Fig 2.5. They may be trying to use multiple Fiber Scanning Displays to reconstruct arbitrary light field. Although the idea is simple and convincing, the light field display is still in their infancy with very high hardware and algorithm complexities, and will not be viable any time soon.

Chapter 3

Depth of Field Image Sequences

The most obvious way to improve image quality of TPVM without more atom frames is to exploit the sparsity of multiview images to be displayed. One example is the sequence of depth of field (DOF) images, which pertaining to continuously varying focal distance but with the position, angle and aperture of the camera fixed. Those images not only provide a strong depth cue, but also have a sparsity structure. It is shown that all member images of the sequence can be approximated with good precision as a linear combination of few basis images. By exploiting the above newly discovered sparsity structure of DOF images, the TPVM multiview display system can support unlimited number of viewers while maintaining an acceptable perceptual quality for all of them, using only a small number of atom frames.

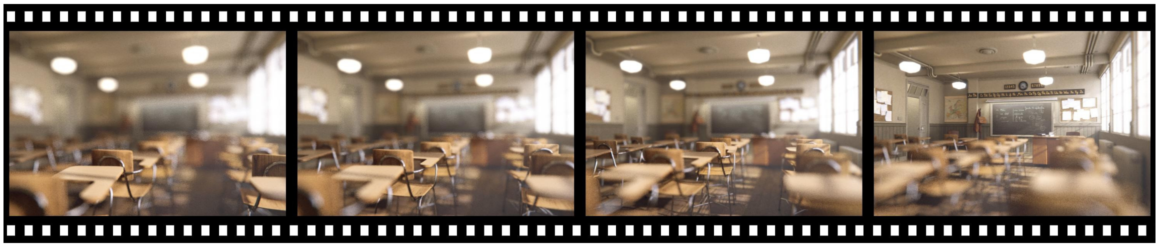


Figure 3.1: An example of DOF image sequence.

3.1 The motivation and definition of DOF image sequences

In 3D computer graphics and virtual reality (VR), a long lasting challenge is to duplicate the percepts of the 3D world on planar displays with maximum realism and minimum visual discomfort. A number of depth cues, including binocular disparity, motion parallax and defocus blur, are used to create the visual sensation of 3D objects on display surfaces. A deep-rooted mainstream technique for 3D perception is stereoscopy that exploits binocular disparity; namely, the display exhibits two different images meant for the left and right eye of a viewer, respectively, for the human visual system (HVS) to interpret the 3D scene from the image differences (Kitamura *et al.*, 2001; Benzie *et al.*, 2007; Love *et al.*, 2009; Geng, 2013).

Meanwhile, another effective and efficient strategy of generating a realistic 3D visual sensation is to combine the depth cues of defocus blur and motion parallax. One way of doing this is real-time gaze-induced depth of the field (DOF) video rendering, with the support of an eye tracking device of sufficient accuracy and speed (Mauderer *et al.*, 2014; Duchowski *et al.*, 2014; Vinnikov and Allison, 2014). Figure 3.1 is an example of gaze-induced DOF sequence, demonstrating how the changes of perspective and focal distance can convincingly convey the sense of depth.

The gaze-induced DOF video is represented by a 4D data set $I(x, y, t, d)$, where (x, y) are pixel coordinates, t the time index, and d the focal distance. Frame t_0 is a sequence of N DOF images $\{I(x, y, t_0, d_n)\}_{1\leq n\leq N}$, which we call a DOF sequence. The sequence $\{I(x, y, t_0, d_n)\}_{1\leq n\leq N}$ consists of the N images taken by a light field camera (real or synthetic) at a fixed position, view angle and time but with different focal distances. Although the data volume of a raw DOF image sequence $\{I(x, y, t_0, d_n)\}_{1\leq n\leq N}$ is N times larger than that of an image or video frame, the underlying 4D signal is highly sparse. Indeed, we will prove that all member images in set $\{I(x, y, t_0, d_n)\}_{1\leq n\leq N}$ can be well approximated as a linear combination of three to four basis images, making it possible to support unlimited number of viewers with TPVM, using only a small number of atom frames.

3.2 Sparsity of DOF image sequences

In this section, we present the analyses to establish the sparsity of the DOF image sequence $\{I(x, y, t_0, d_n)\}_{1\leq n\leq N}$ consisting of images of the same scene captured or rendered with N different focal distance settings. Specifically, we show that three to four basis images suffice to linearly approximate all member images in a DOF image sequence accurately, regardless of the size of the sequence, N. For the sake of simplicity, we omit pixel coordinates x, y and frame index t_0 in $I(x, y, t_0, d)$ and use $I(d)$ to denote a DOF image in the following discussions.

Suppose $I(d)$ is an image of a single flat object placed at distance C_D from the camera; and the focal distance, focal length and aperture size of the camera are d, C_F, C_A, respectively. If f is a focused image of the object, i.e., $f = I(C_D)$, then the defocus blur effect due to large aperture can be simulated using image convolution as follows,

$$I(d) = f * g(d) \tag{3.1}$$

where $g(d)$ is a blur kernel. The exact shape of the blur kernel is camera specific, but by the basic optical properties of camera lens, the support size of kernel $g(d)$ can be calculated as,

$$s = \left| C_A \frac{C_F(d - C_D)}{C_D(d - C_F)} \right|, \tag{3.2}$$

which is the size of the confusion circle.

Now the question is that, given a DOF image sequence $\{I(d_n)\}_{1 \le n \le N}$, if there exists a small number of basis $\Phi = \{\phi_1, \phi_2, \ldots, \phi_M\}$ that can reconstruct the sequence with sufficient accuracy. This problem can be formulated as an unconstrained optimization problem, minimizing the reconstruction error of basis variable Φ and coefficient vector variable $w(d_n)$ for each DOF image $I(d_n)$,

$$\min_{\Phi, w} \frac{1}{NL} \sum_{n=1}^{N} \| I(d_n) - \Phi w(d_n) \|_2^2, \tag{3.3}$$

where L is the number of pixels in each DOF image. As both Φ and w are variables, the problem in Eq. (3.3) is non-convex in this form. However, since there is no constraint on either Φ or w, this problem is equivalent to finding a low-rank approximation of matrix $I = \{I(d_n)\}$ where each column vector is a DOF image,

$$\begin{aligned} \min_{\tilde{I}} \quad & \|I - \tilde{I}\|_{\mathrm{F}} \\ \text{s.t.} \quad & \operatorname{rank}(\tilde{H}) \leq M. \end{aligned} \tag{3.4}$$

This problem is tractable using singular value decomposition (SVD). Suppose the SVD of H is $U\Sigma V^{\intercal}$, and diagonal matrix Σ' preserves only the first M singular values in Σ. Then $\Phi = U\Sigma'$ and $w = V^{\intercal}$ is an optimal solution of the original problem.

The optimal solution of Eq. (3.3) is the minimum error of reconstructed sequence using a small number of basis images. If the optimal solution is bounded for any DOF image sequence, then there exists a good basis with low reconstruction error for each DOF image sequence. However, the optimal solution is difficult to estimate directly, as it depends on not only the characteristics of the camera, i.e., blur kernel $g(d_n)$, but also the appearance of object, i.e., the focused image f. These two factors, however, can be separated easily in frequency domain by convolution theorem. Let $\mathcal{F}$ be the Fourier transform matrix. Since matrix $\mathcal{F}$ is unitary, the original optimization problem is equivalent to

$$\begin{aligned} &\min_{\Phi,w} \frac{1}{NL} \sum_{n=1}^{N} \|\mathcal{F}\|_2^2 \cdot \|I(d_n) - \Phi w(d_n)\|_2^2 \\ =&\min_{\Phi,w} \frac{1}{NL} \sum_{n=1}^{N} \|\mathcal{F}I(d_n) - \mathcal{F}\Phi w(d_n)\|_2^2 \\ =&\min_{\Phi,w} \frac{1}{NL} \sum_{n=1}^{N} \|F \circ G(d_n) - \mathcal{F}\Phi w(d_n)\|_2^2, \end{aligned} \tag{3.5}$$

where vectors F and $G(d_n)$ are the Fourier transform of f and $g(d_n)$, respectively, and vector $F \circ G(d_n)$ represents the piecewise product of F and $G(d_n)$. For some optimal solution Φ, w, these must exist $\tilde{G}(d_n)$, such that,

$$F \circ \hat{G}(d_n) = \hat{\Phi} w(d_n), \tag{3.6}$$

where the i-th row of basis matrix $\hat{\Phi}$ is

$$\hat{\Phi}_i = \begin{cases} 0 & \text{if } F_i = 0 \\ [\mathcal{F}\Phi]_i & \text{if } F_i \neq 0. \end{cases} \tag{3.7}$$

The idea is that, by setting row $\hat{\Phi}_i$ to be zero, the reconstruction error of the i-th element of $F \circ G(d_n)$ is always zero without affecting the optimality of the other rows. Thus, if solution Φ, w is optimal, $\mathcal{F}^{-1}\hat{\Phi}, w$ must also be optimal for the problem in Eq. (3.5). Therefore, this problem is equivalent to

$$\begin{aligned} \min_{\hat{G},\hat{\Phi},w} \quad & \frac{1}{NL}\sum_{n=1}^{N} \|F \circ G(d_n) - F \circ \hat{G}(d_n)\|_2^2, \\ \text{s.t.} \quad & F \circ \hat{G}(d_n) = \hat{\Phi} w(d_n). \end{aligned} \tag{3.8}$$

Since a low-rank approximation of matrix G, say $\tilde{G}$, might be not optimal as a solution of variable $\hat{G}$ in Eq. (3.8), the optimal solution of the problem must be less than or equal to,

$$
\begin{aligned}
&\frac{1}{NL}\sum_{n=1}^{N}\|F\circ[G(d_n)-\tilde{G}(d_n)]\|_2^2\\
=&\frac{1}{NL}\sum_{i=1}^{L}F_i^2\cdot\|G_i-\tilde{G}_i\|_2^2\\
\le&\frac{1}{NL}\left(\sum_{i=1}^{L}F_i^2\right)\cdot\left(\max_{i=1}^{L}\|G_i-\tilde{G}_i\|_2^2\right)\\
=&\left(\frac{1}{L}\|f\|_2^2\right)\cdot\left(\frac{1}{N}\max_{i=1}^{L}\|G_i-\tilde{G}_i\|_2^2\right),
\end{aligned}
\tag{3.9}
$$

where and $G_i, \tilde{G}_i$ are the i-th row of $G, \tilde{G}$, respectively. Suppose blur kernel $g(d)$ is Gaussian, which is commonly used in DOF rendering,

$$g(d)=\frac{\sqrt{\pi}}{s}e^{-\frac{\pi^2k^2}{s^2}}, \tag{3.10}$$

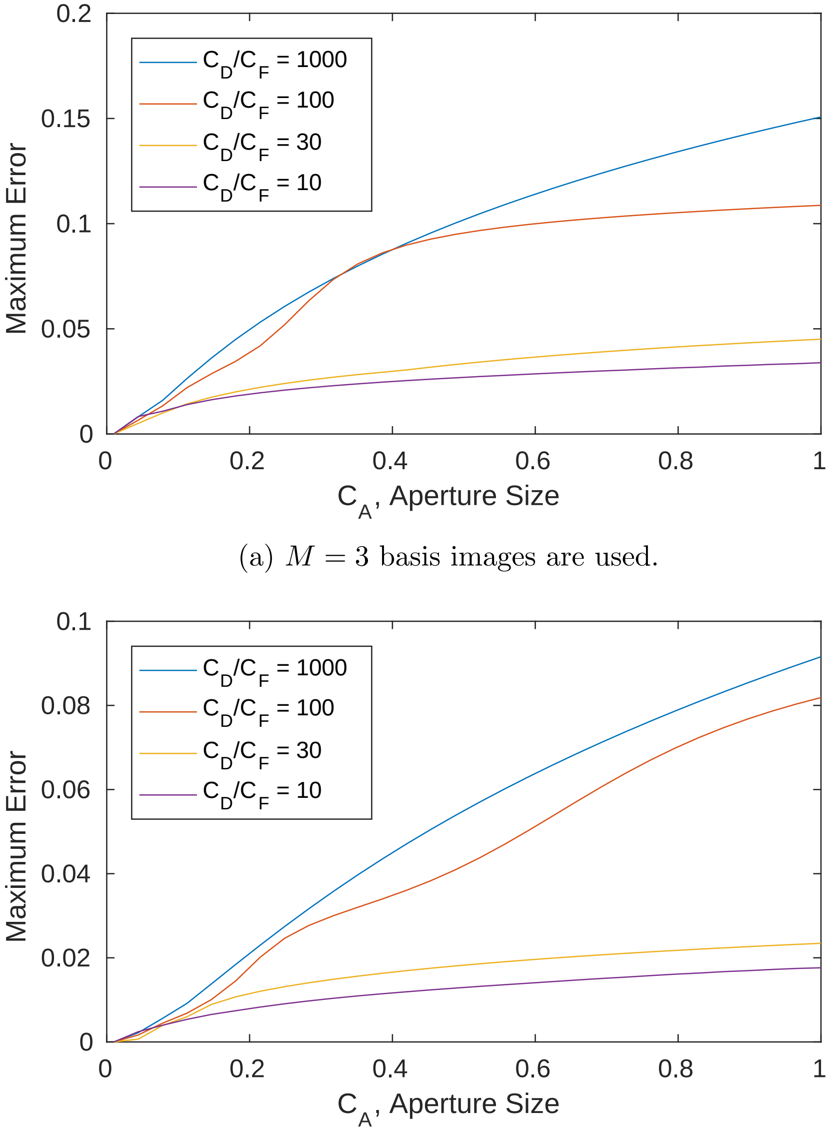
then its frequency domain image $G(d)$ is also Gaussian,

$$G(d)=e^{-s^2x^2}. \tag{3.11}$$

As plotted in Figure 3.2, numerical results show that, if focal distance d_n changes from $10C_F$ to $1000C_F$ in a DOF image sequence of $N = 1000$ images, then the maximum root mean square error (RMSE) of row vectors of $G-\tilde{G}$, i.e.,

$$\max_{i=1}^{L}\sqrt{\frac{1}{N}\|G_i-\tilde{G}_i\|_2^2}, \tag{3.12}$$

is less than 0.15 when $M = 3$ basis images are used, and less than 0.09 when $M = 4$ basis images are used. By the error bound inequality in Eq. (3.9), the RMSE of



(a) $M = 3$ basis images are used.

(b) $M = 4$ basis images are used.

Figure 3.2: The maximum root mean square error of row vectors of $G - \tilde{G}$ if focal distance d_n changes from $10C_F$ to $1000C_F$ in a DOF image sequence of $N = 1000$ images.

the reconstructed DOF image sequence by using Eq. 3.3 is less than $0.09 \cdot \|f\|_2/\sqrt{L}$ for $M = 4$, implying that the relative error of the reconstructed images is less than 9% roughly. The case when $M = 3$ is similar. This error bound only covers the reconstruction error of Eq. (3.3) in the worst case scenario; for the majority cases we tested, the reconstruction error is far below the bound.

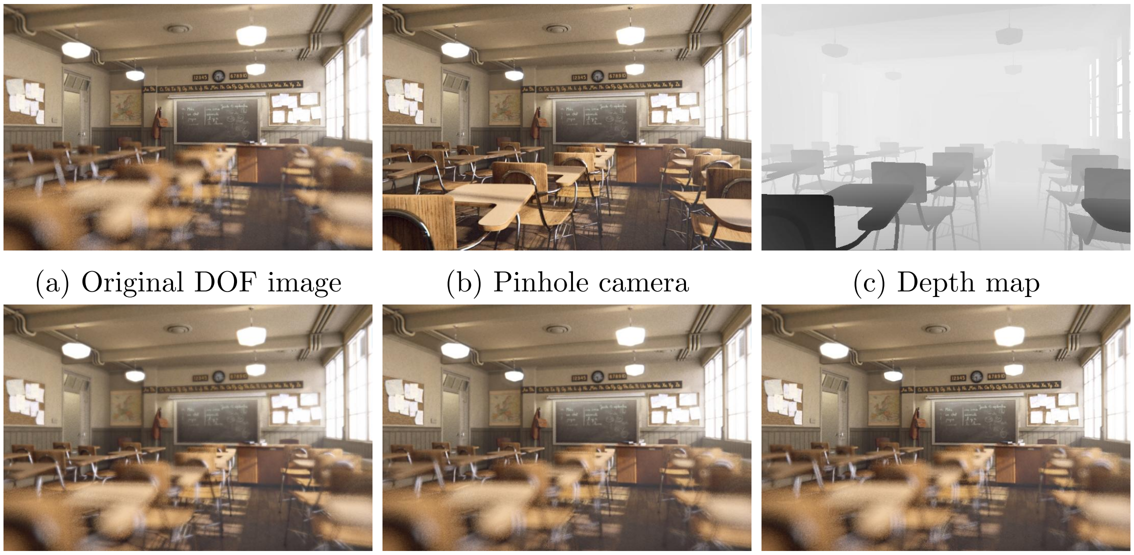
So far, we have already proved that any DOF image sequence can be well approximated as a linear combination of three to four basis images. However, there is one caveat. The image decomposition underlying TPVM has to respect a condition of non-negativity, because the light energy emitted by the display cannot be negative, and active LC glasses can only implement modulation weights between 0 and 1. Therefore, the proof cannot completely apply to the TPVM multiview display system. But the good news is that the non-negativity constraints have almost no impact on the reconstruction error in our simulation. We are confident that most of DOF image sequences can be reconstructed well in TPVM, using only a small number of atom frames.

3.3 Experimental results

In this section we report our experimental results. We use the projection to convex set (POCS) method (Bauschke and Borwein, 1996) to solve the NMF problem of (2.1). In the POCS approach, the nonnegativity constraints when alternatingly solving (2.1) for $\boldsymbol{W}$ and $\boldsymbol{X}$ are relaxed; but after each iteration, the resulting matrices $\boldsymbol{W}$ and $\boldsymbol{X}$ are clipped back to the value range $[0, 1]$. This allows the NMF problem to be solved very efficiently as a series of least-squares problems.

The test DOF image sequences are generated by Blender (Blender, 2017), an open source 3D creation suite. Listed in Figures 3.3, 3.4, 3.5 and 3.6 are the rendered DOF results of two 3D demo models, “Class Room” and “Barcelona Pavilions”, which are available on the Blender website. For a fixed camera position we rendered a series of DOF images with varying focal distances from the nearest to the farthest objects and use them as reference DOF images (ground truth).

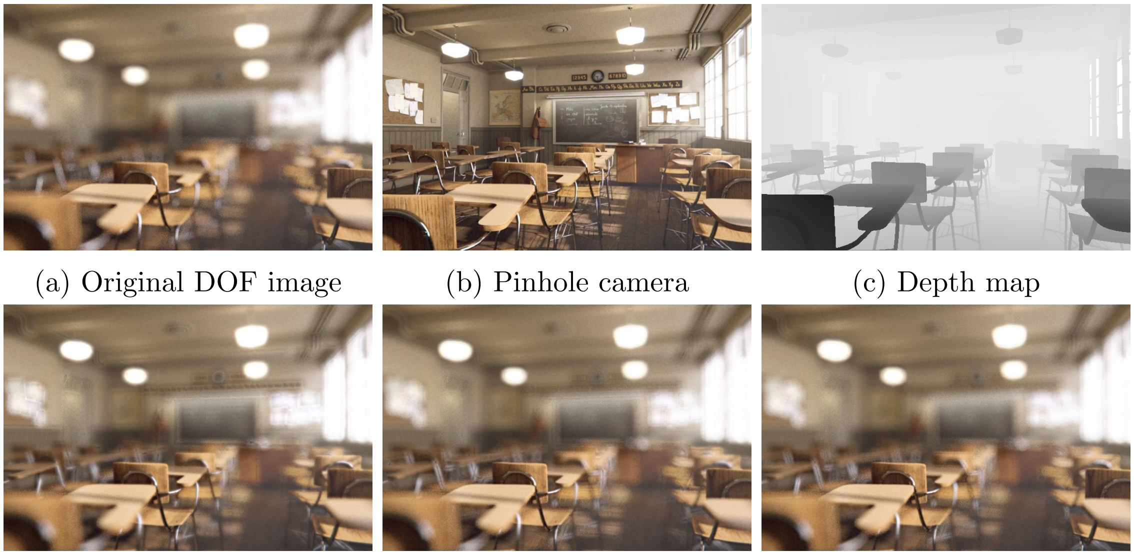
We also test our new method with DOF sequence captured by Lytro Illum camera, a real light field camera which can capture a DOF sequence in only one exposure. The results are shown in Figures 3.7. We find that both synthetic and real DOF image sequences can be reconstructed well in TPVM, using only a small number of atom frames.



(a) Original DOF image (b) Pinhole camera (c) Depth map

(d) Our method with 2 bases (30.91 dB) (e) Our method with 3 bases (35.27 dB) (f) Our method with 4 bases (40.70 dB)

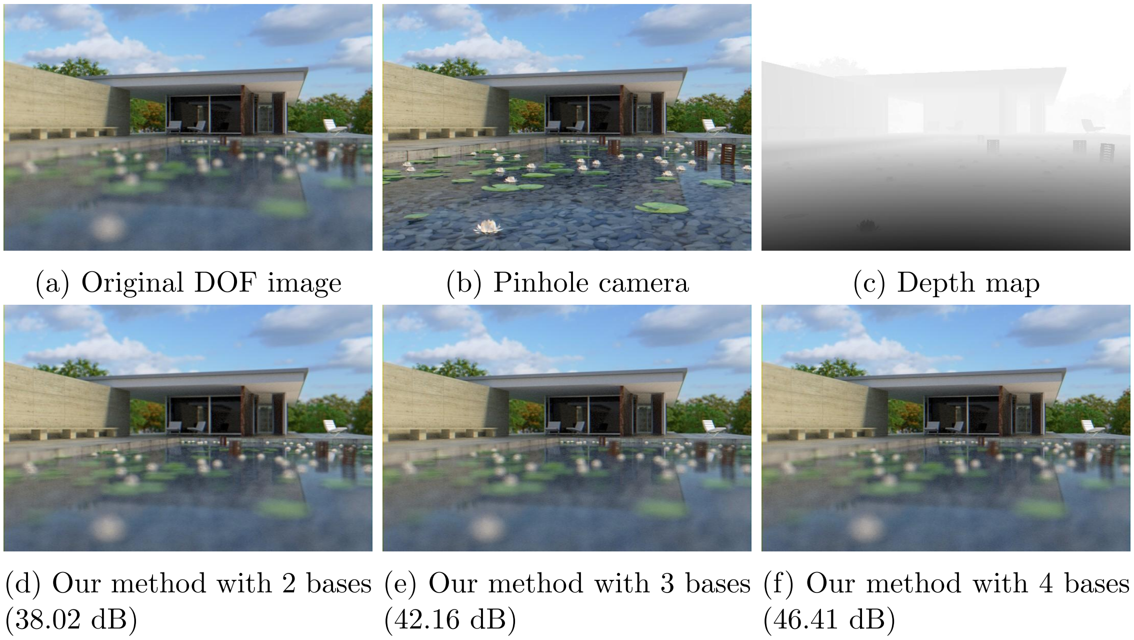
Figure 3.3: TPVM outputs of DOF sequence in classroom scene with far focus.



(a) Original DOF image (b) Pinhole camera (c) Depth map

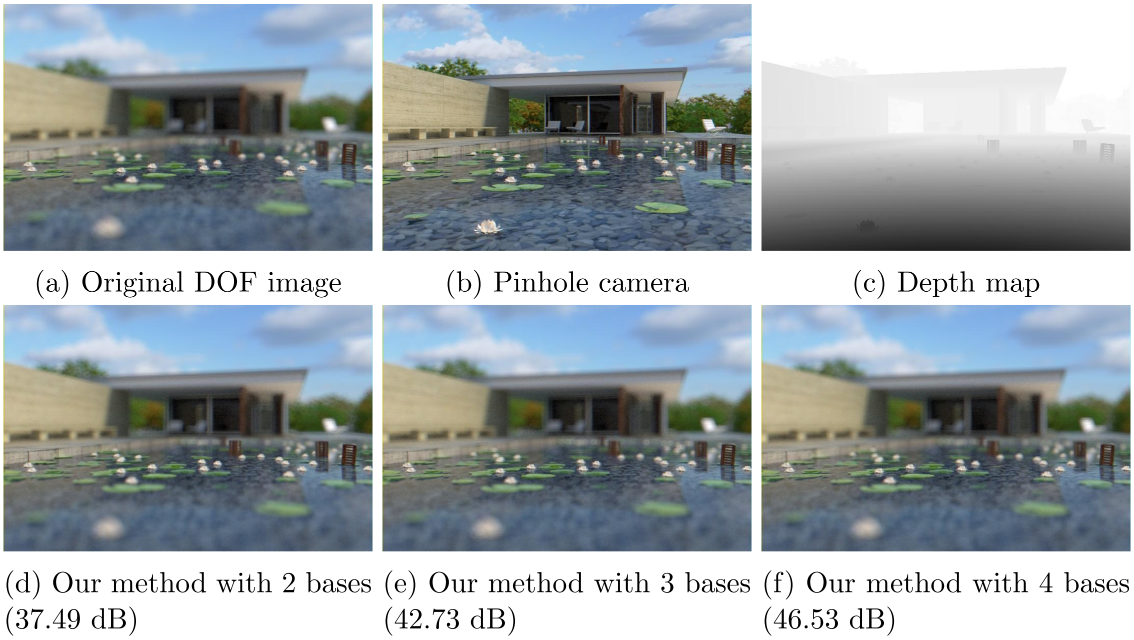
(d) Our method with 2 bases (31.88 dB) (e) Our method with 3 bases (36.56 dB) (f) Our method with 4 bases (39.37 dB)

Figure 3.4: TPVM outputs of DOF sequence in classroom scene with near focus.


(a) Original DOF image (b) Pinhole camera (c) Depth map

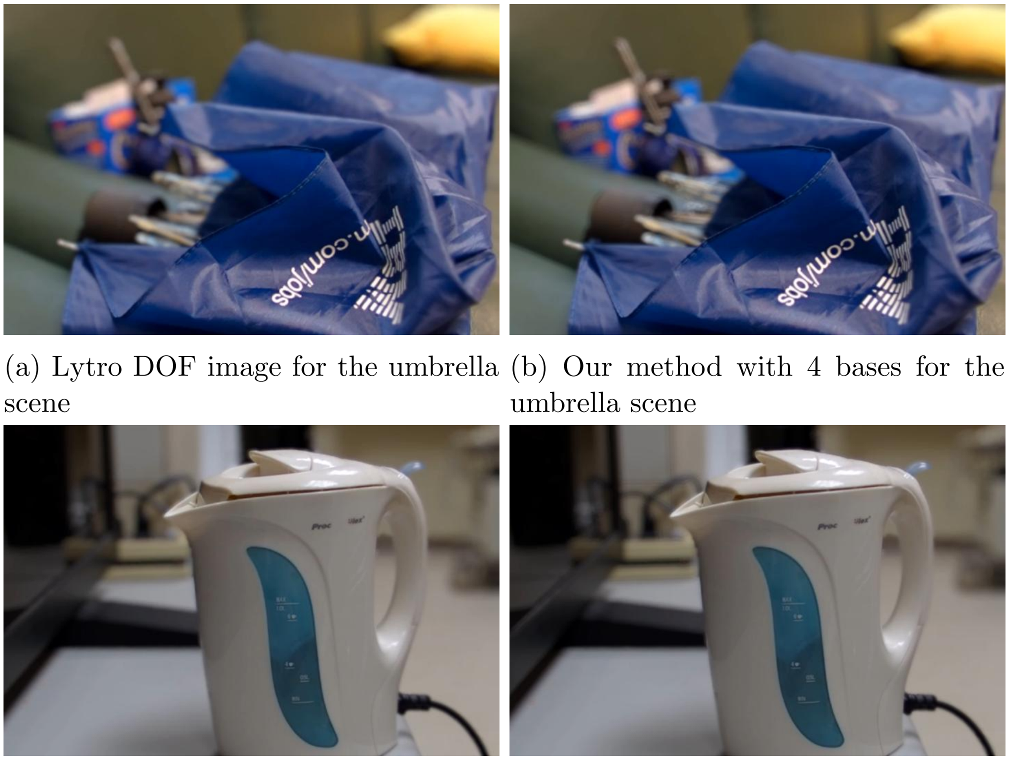
(d) Our method with 2 bases (38.02 dB) (e) Our method with 3 bases (42.16 dB) (f) Our method with 4 bases (46.41 dB)

Figure 3.5: TPVM outputs of DOF sequence in the pavilion scene with far focus.


(a) Original DOF image (b) Pinhole camera (c) Depth map

(d) Our method with 2 bases (37.49 dB) (e) Our method with 3 bases (42.73 dB) (f) Our method with 4 bases (46.53 dB)

Figure 3.6: TPVM outputs of DOF sequence in the pavilion scene with near focus.



(a) Lytro DOF image for the umbrella scene

(b) Our method with 4 bases for the umbrella scene

(c) Lytro DOF image for the kettle scene

(d) Our method with 4 bases for the kettle scene

Figure 3.7: TPVM outputs of DOF sequence captured by Lytro Illum camera.

Chapter 4

Fovea Weighting of TPVM

In many applications of multiview displays, different users often have their own regions of interest at any given time. This allows us to exploit the well-known property of rapidly decreasing visual acuity from fovea to peripheral vision (Weymouth, 1958), and propose a spatially weighted optimization algorithm for multiview computational display. The proposed algorithm chooses the basis images and their fusion scheme in such a way that different concurrent views are exhibited at highest quality in viewers focused regions, while allowing graceful image quality degradation in regions of peripheral vision. In this chapter, we discuss how to apply fovea weighting in the framework of TPVM; namely, formulate a weighted objective function and devise efficient algorithms to minimize it. We find that viewers' perceptual quality improved significantly in this way, especially when the TPVM multiview display system only has a small number of atom frames.

4.1 Fovea weighting

Assuming that the users of the TPVM multiview display system are all equipped with eye tracking devices of sufficiently high accuracy and response speed (Duchowski, 2007; Holmqvist *et al.*, 2011; Mauderer *et al.*, 2014), thus the region of interests (ROI) for each viewer is known. For viewer k, the approximation error image $\boldsymbol{y}_k - \boldsymbol{X}\boldsymbol{w}_k$ is spatially weighted by a 2D Gaussian function that is aligned with the ROI of viewer k.

Let $\boldsymbol{C}$ be the K corresponding regions of interest of K target images. The $N \times K$ non-negative matrix $\boldsymbol{C}$, which has the same size as $\boldsymbol{Y}$, can be used as a weight matrix inside the Frobenius norm. Now we need to solve the following optimization problem with ROI information:

$$\min_{\boldsymbol{X},\boldsymbol{W}} \|\boldsymbol{C} \circ (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{W})\|_F^2, \quad \text{subject to} \quad 0 \leq \boldsymbol{X}, \boldsymbol{W} \leq 1 \tag{4.1}$$

where $\circ$ is the Hadamard product. First, we suppose $\boldsymbol{W}$ is the only optimization variable in our problem, and the objective is to minimize

$$\begin{aligned} S(\boldsymbol{W}) &= \sum_i \|\boldsymbol{c}_i \circ (\boldsymbol{y}_i - \boldsymbol{X}\boldsymbol{w}_i)\|^2 \\ &= \sum_i (\boldsymbol{c}_i \circ (\boldsymbol{y}_i - \boldsymbol{X}\boldsymbol{w}_i))^T (\boldsymbol{c}_i \circ (\boldsymbol{y}_i - \boldsymbol{X}\boldsymbol{w}_i)) \end{aligned} \tag{4.2}$$

where $\boldsymbol{w}_i$ refers to the ith column of $\boldsymbol{W}$. Let diagonal matrix $\boldsymbol{D}_i = diag(\boldsymbol{c}_i)$, and the objective is to minimize

$$\begin{aligned} S(\boldsymbol{W}) =& \sum_i (\boldsymbol{y}_i^T \boldsymbol{D}_i - \boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i)(\boldsymbol{D}_i \boldsymbol{y}_i - \boldsymbol{D}_i \boldsymbol{X} \boldsymbol{w}_i) \\ =& \sum_i (\boldsymbol{y}_i^T \boldsymbol{D}_i^2 \boldsymbol{y}_i - \boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{y}_i - \boldsymbol{y}_i^T \boldsymbol{D}_i^2 \boldsymbol{X} \boldsymbol{w}_i + \boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{X} \boldsymbol{w}_i) \end{aligned} \tag{4.3}$$

Note that: $(\boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{y}_i)^T = \boldsymbol{y}_i^T \boldsymbol{D}_i^2 \boldsymbol{X} \boldsymbol{w}_i$ is a scalar and equal to its own transpose, hence $\boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{y}_i = \boldsymbol{y}_i^T \boldsymbol{D}_i^2 \boldsymbol{X} \boldsymbol{w}_i$ and the quantity to minimize becomes

$$S(\boldsymbol{W}) = \sum_i (\boldsymbol{y}_i^T \boldsymbol{D}_i^2 \boldsymbol{y}_i - 2\boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{y}_i + \boldsymbol{w}_i^T \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{X} \boldsymbol{w}_i) \tag{4.4}$$

Differentiating this with respect to $\boldsymbol{w}_i$ and equating to zero to satisfy the first-order conditions gives

$$\begin{aligned} \frac{\partial S}{\partial \boldsymbol{w}_i} =& -2\boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{y}_i + 2(\boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{X}) \boldsymbol{w}_i \\ =& -2\boldsymbol{X}^T diag^2(\boldsymbol{c}_i) \boldsymbol{y}_i + 2(\boldsymbol{X}^T diag^2(\boldsymbol{c}_i) \boldsymbol{X}) \boldsymbol{w}_i \\ =& 0 \end{aligned} \tag{4.5}$$

Suppose $\boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{X}$ is positive definite, then we can get solution for $\boldsymbol{w}_i$

$$\begin{aligned} \boldsymbol{w}_i =& (\boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{D}_i^2 \boldsymbol{y}_i \\ =& (\boldsymbol{X}^T diag^2(\boldsymbol{c}_i) \boldsymbol{X})^{-1} \boldsymbol{X}^T diag^2(\boldsymbol{c}_i) \boldsymbol{y}_i \end{aligned} \tag{4.6}$$

Suppose $\boldsymbol{X}$ is the only optimization variable in our problem, we can get the partial derivatives and the solution for $\boldsymbol{x}_i^T$ in a similar way.

$$\frac{\partial S}{\partial \boldsymbol{x}_i^T} = -2\boldsymbol{y}_i^T diag^2(\boldsymbol{c}_i^T)\boldsymbol{W}^T + 2\boldsymbol{x}_i^T(\boldsymbol{W} diag^2(\boldsymbol{c}_i^T)\boldsymbol{W}^T) \tag{4.7}$$

$$\boldsymbol{x}_i^T = \boldsymbol{y}_i^T diag^2(\boldsymbol{c}_i^T)\boldsymbol{W}^T(\boldsymbol{W} diag^2(\boldsymbol{c}_i^T)\boldsymbol{W}^T)^{-1} \tag{4.8}$$

where $\boldsymbol{x}_i^T$ refers to the ith row of $\boldsymbol{X}$.

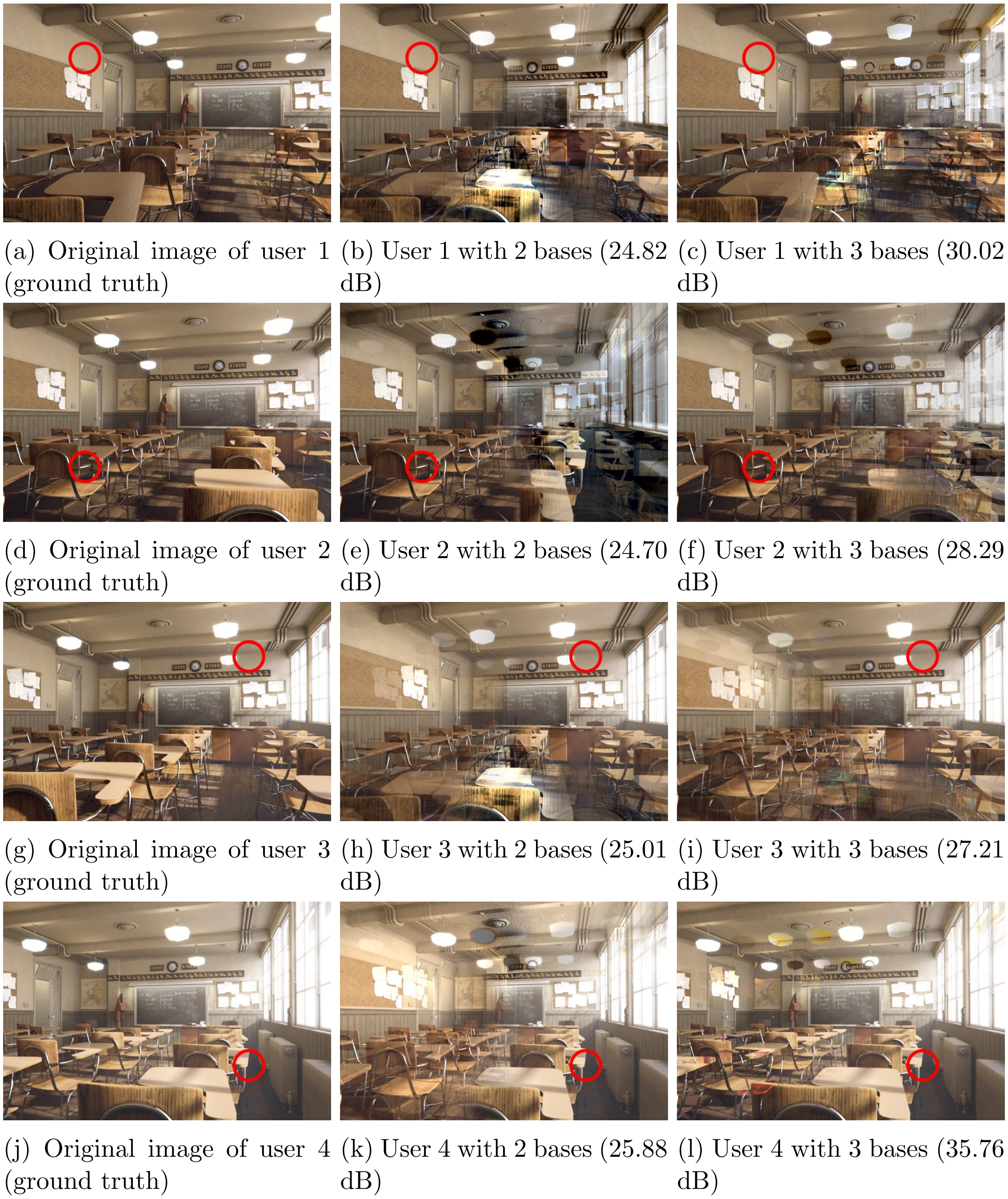
It is worth noting that $diag(\boldsymbol{c}_i)$ has a huge size in our problem. To accelerate the computation of $\boldsymbol{W}$, we can replace the usage of diagonal matrix with Hadamard product and tiling of copies of $\boldsymbol{c}_i$. For example, compute $(\boldsymbol{c}_i, \boldsymbol{c}_i, \cdots, \boldsymbol{c}_i) \circ \boldsymbol{X}$ in the program instead of $diag(\boldsymbol{c}_i)\boldsymbol{X}$.

We suggest to use the projected gradient descent method (Lin, 2007) to solve the weighted NMF problem, i.e., perform usual gradient update and then project back onto the convex feasible set. The computation of the gradient needed in this method is provided in (4.5) and (4.7). The projected gradient descent method can obtain a better convergence if a good starting value is chosen. So in practice, we can use the result of POCS method (Bauschke and Borwein, 1996) as the starting value of the projected gradient descent method.

4.2 Experimental results

In the experimental setup, we try to accommodate the hardware limitation of the TPVM multiview display system in practice. Although the number of basis images, can be any positive integer theoretically, its feasible range is severely limited by the relatively low speed of active LC light modulation glasses. The speed of the off-the-shelf LC glasses of grey levels is difficult to exceed 180Hz. In other words, at the current state of the art, a practical TPVM multiview display system can only run two to three basis images. As such, as said in the very beginning of the thesis, the main technical challenge facing TPVM is how to support as many viewers as possible while maintaining an acceptable perceptual quality for all of them, using only a small number of basis images. In the experiment reported below, we demonstrate that fovea weighting can achieve high perceptual quality for four different concurrent views with only two or three basis images, which is impossible with naive temporal multiplexing as in the main stream stereoscopic displays.

The test images in our experiment are generated by Blender (Blender, 2017), an open source 3D creation suite. Listed images in Figure 4.1 and 4.2 are the rendered images of two 3D demo models, “Class Room” and “Barcelona Pavilions”, which are available on the Blender website. In the experiment, four users are watching the classroom with four different perspectives and regions of interest. The red circles in Figure 4.1 and 4.2 indicate the centers of regions of interest of different users. Fovea weighted PSNRs are also reported in figures.



(a) Original image of user 1 (ground truth)

(b) User 1 with 2 bases (24.82 dB)

(c) User 1 with 3 bases (30.02 dB)

(d) Original image of user 2 (ground truth)

(e) User 2 with 2 bases (24.70 dB)

(f) User 2 with 3 bases (28.29 dB)

(g) Original image of user 3 (ground truth)

(h) User 3 with 2 bases (25.01 dB)

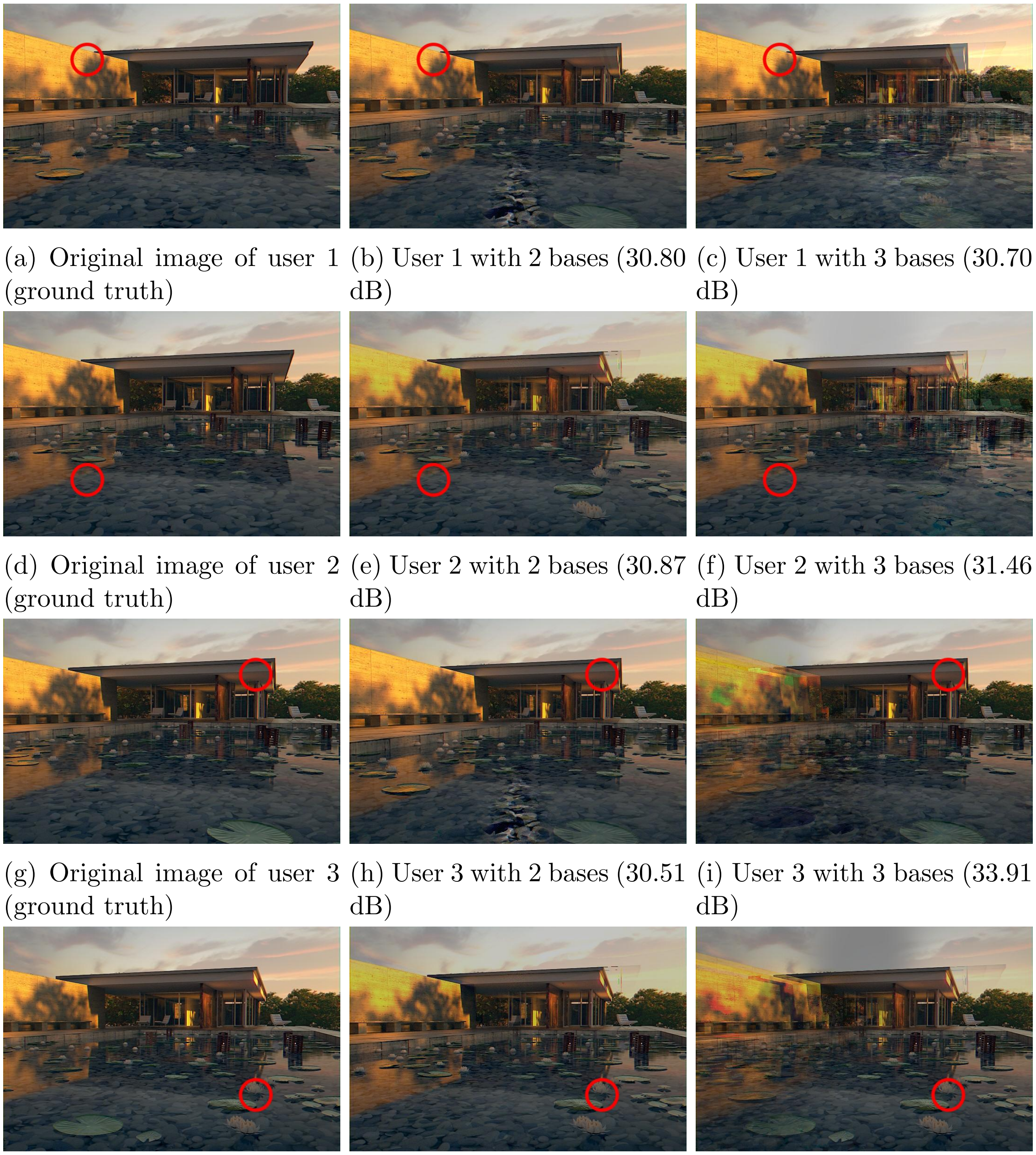
(i) User 3 with 3 bases (27.21 dB)

(j) Original image of user 4 (ground truth)

(k) User 4 with 2 bases (25.88 dB)

(l) User 4 with 3 bases (35.76 dB)

Figure 4.1: Results of fovea weighting of TPVM in classroom scene. The red circles indicate the centers of regions of interest of different users. The PSNRs reported are fovea weighted PSNR.



(a) Original image of user 1 (ground truth)

(b) User 1 with 2 bases (30.80 dB)

(c) User 1 with 3 bases (30.70 dB)

(d) Original image of user 2 (ground truth)

(e) User 2 with 2 bases (30.87 dB)

(f) User 2 with 3 bases (31.46 dB)

(g) Original image of user 3 (ground truth)

(h) User 3 with 2 bases (30.51 dB)

(i) User 3 with 3 bases (33.91 dB)

(j) Original image of user 4 (ground truth)

(k) User 4 with 2 bases (31.75 dB)

(l) User 4 with 3 bases (31.29 dB)

Figure 4.2: Results of fovea weighting of TPVM in the pavilion scene. The red circles indicate the centers of regions of interest of different users. The PSNRs reported are fovea weighted PSNR.

Chapter 5

Conclusions

Compared with other computational displays, TPVM is more feasible, more suitable for multiview applications, and has a brighter, sharper and cleaner image. Without a quicker alternative to current liquid crystal glasses, our study has great realistic significance for us to build a practical TPVM multiview display system. In this study, we develop two approaches to achieve an acceptable perceptual quality for all of viewers, with the off-the-shelf but still relatively slow LC glasses.

The first approach is to exploit the sparsity of multiview images to be displayed. One example is the sequence of depth of field (DOF) images, which pertaining to continuously varying focal distance but with the position, angle and aperture of the camera fixed. Those images not only provide a strong depth cue, but also have a sparsity structure. That structure allows the DOF images to be reconstructed from a small number of atom frames. We prove that property theoretically and experimental results agree well with the proof.

The second approach is to exploit the well-known property of rapidly decreasing visual acuity from fovea to peripheral vision. We propose a spatially weighted optimization algorithm for TPVM based on viewers' real time region of interests (ROI) information. The strategy is to exhibit different concurrent views at highest quality in viewers' focused regions, while allowing graceful image quality degradation in regions of peripheral vision. This is achieved by a novel fovea weighting algorithm that optimizes for subjective quality. We find the proposed algorithm improves viewers' perceptual quality significantly, especially when the TPVM multiview display system only has a small number of atom frames.

The effectiveness of both methods is validated by simulation results, and both ways can reach the same goal of improving perceptual quality for TPVM, using only a small number of atom frames.

Bibliography

Bauschke, H. H. and Borwein, J. M. (1996). On projection algorithms for solving convex feasibility problems. *SIAM review*, **38**(3), 367–426.

Benzie, P., Watson, J., Surman, P., Rakkolainen, I., Hopf, K., Urey, H., Sainov, V., and von Kopylow, C. (2007). A survey of 3DTV displays: techniques and technologies. *IEEE Transactions on Circuits and Systems for Video Technology*, **17**(11), 1647–1658.

Blender (2017). blender.org - home of the blender project - free and open 3d creation software. Available at: https://www.blender.org/.

Crossman-Bosworth, J., Seibel, E. J., and Fauver, M. E. (2006). Optical beam scanning system for compact image display or image acquisition. US Patent 7,068,878.

Duchowski, A. T. (2007). Eye tracking methodology. *Theory and practice*, **328**.

Duchowski, A. T., House, D. H., Gestring, J., Wang, R. I., Krejtz, K., Krejtz, I., Mantiuk, R., and Bazyluk, B. (2014). Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the ACM Symposium on Applied Perception*, pages 39–46. ACM.

Geng, J. (2013). Three-dimensional display technologies. *Advances in optics and photonics*, **5**(4), 456–535.

Google (2014). Google glass. Available at: https://www.x.company/glass/.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

HTC (2016). Vive — discover virtual reality beyond imagination. Available at: https://www.vive.com/eu/.

Kitamura, Y., Konishi, T., Yamamoto, S., and Kishino, F. (2001). Interactive stereoscopic display for three or more users. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 231–240. ACM.

Kramida, G. (2016). Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, **22**(7), 1912–1931.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**(6755), 788–791.

Lee, D. D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562.

Lin, C.-J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural computation*, **19**(10), 2756–2779.

Love, G. D., Hoffman, D. M., Hands, P. J., Gao, J., Kirby, A. K., and Banks, M. S. (2009). High-speed switchable lens enables the development of a volumetric stereoscopic display. *Optics express*, **17**(18), 15716–15725.

Masia, B., Wetzstein, G., Didyk, P., and Gutierrez, D. (2013). A survey on computational displays: Pushing the boundaries of optics, computation, and perception. *Computers & Graphics*, **37**(8), 1012–1038.

Mauderer, M., Conte, S., Nacenta, M. A., and Vishwanath, D. (2014). Depth perception with gaze-contingent depth of field. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 217–226. ACM.

Microsoft (2016). Microsoft hololens — transform your world with holograms. Available at: https://www.microsoft.com/en-ca/hololens.

Oculus (2016). Oculus rift, a virtual reality system that completely immerses you inside virtual worlds. Available at: https://www.oculus.com/rift/.

Samsung (2015). Samsung gear vr with controller. Available at: http://www.samsung.com/global/galaxy/gear-vr/.

Schowengerdt, B. T., Johnston, R. S., Lee, C. M., Melville, C. D., and Seibel, E. J. (2010). 1 mm x 7 mm full-color pico projector using scanning optical fiber. In *17th International Display Workshops, IDW'10*.

Sony (2016). Playstation vr an upcoming virtual reality headset for ps4 console. Available at: https://www.playstation.com/en-ca/explore/playstation-vr/.

Vinnikov, M. and Allison, R. S. (2014). Gaze-contingent depth of field in realistic scenes: The user experience. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 119–126. ACM.

Wetzstein, G., Lanman, D. R., Hirsch, M. W., and Raskar, R. (2012). Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. *ACM Transaction on Graphics (ToG)*, **31**(4), 80:1–80:11.

Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American journal of ophthalmology*, **46**(1), 102–113.

Wu, X. and Zhai, G. (2013). Temporal psychovisual modulation: A new paradigm of information display [exploratory dsp]. *IEEE Signal Processing Magazine*, **30**(1), 136–141.

Zhai, G. and Wu, X. (2014). Multiuser collaborative viewport via temporal psychovisual modulation. *IEEE Signal Processing Magazine*, **31**(5), 144–149.