TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

UGUR AKPINAR

LENS ARRAY BASED TECHNIQUES FOR 3D SCENE CAPTURE AND DISPLAY

Master of Science Thesis

Examiners: Prof. Atanas Gotchev, Dr. Erdem Sahin
Examiner and topic approved by the Faculty Council of the Faculty of Computing and Electrical Engineering on 20th September 2017

ABSTRACT

UGUR AKPINAR: Lens Array Based Techniques for 3D Scene Capture and Display
Tampere University of Technology
Master of Science Thesis, 48 pages, 0 Appendix pages
September 2017
Master's Degree Programme in Information Technology
Major: Signal Processing
Examiners: Prof. Atanas Gotchev, Dr. Erdem Sahin
Keywords:

This thesis discusses the use of lens arrays for both capture and display of 3D visual scenes while utilizing the ray optics formalism for modeling the propagation of light. In 3D capture, the use of lens arrays brings the concepts of focused and defocused plenoptic cameras, and in 3D display, the same optical technology brings the integral imaging (InI) and super multiview (SMV) visualization techniques.

Plenoptic cameras combine a lens array with a single sensor in order to capture the light field (LF) emanated by a scene compactly and in a single shot. In the thesis, comparative analysis of focused and defocused plenoptic cameras is carried out in terms of LF sampling and spatio-angular resolution trade-offs. An algorithm for simulating ground-truth plenoptic image data for the case of defocused plenoptic camera is developed and implemented. It models the process of plenoptic capture and makes use of the notion of densely sampled light field (DSLF) for the sake of efficient and reliable data processing.

3D displays aim at visualising 3D scenes as accurate as possible, thus providing natural viewing experience. They are characterised and compared by their ability to correctly reproduce 3D visual cues, such as vergence, binocular disparity, accommodation and motion parallax. Design-wise, lens array based 3D display techniques provide simple yet effective way to correctly deliver all these cues, which makes them attractive in several 3D display applications. The thesis studies SMV and InI techniques in terms of depth perception and resolution trade-offs. Based on the theoretical analysis, a prototype SMV head-up display (HUD) system is developed. It demonstrates a compact and affordable solution for the virtual image presentation HUD problem. The experiments and analyses carried out on the prototype verify the SMV display capabilities for the targeted HUD application.

PREFACE

This thesis work is completed with the 3D Media group in the Laboratory of Signal Processing, Tampere University of Technology. Ray optics formalism is applied to both 3D display and capture problems.

I would like to thank Prof. Atanas Gotchev for his supervision during my thesis. Besides, I would like to thank Dr. Erdem Sahin for supporting and motivating me all along the path. It is my pleasure to declare that during my work in the 3D Media group, I experienced a delightful environment with all my colleagues. Therefore, I feel myself lucky to build my future with such creative people.

During my studies, I gained hands-on experience in the field and broadened my vision. In this aspect, the Laboratory of Signal Processing has an unforgettable effect on me. I appreciate all my teachers for their effort.

Last but not least, I am thankful to my family that they have never withdrawn their support. Thanks for your understanding and being with me all the way to my graduation. I love you.

Tampere, 20.9.2017

Ugur Akpinar

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

3D	Three-dimensional
AR	Augmented reality
CGR	Computer graphics rendering
DoF	Depth of field
DSLF	Densely sampled light field
EPI	Epipolar plane image
FOV	Field of view
HMD	Head-mounted display
HPO	Horizontal-parallax-only
HUD	Head-up display
IBR	Image based rendering
InI	Integral Imaging
LCD	Liquid crystal display
LF	Light field
MV	Multiview
SLM	Spatial light modulator
SMV	Super Multiview

a	distance from the microlenses to the image plane in the focused plenoptic camera
B	baseline between adjacent views
b	distance from the sensor to microlenses in the focused plenoptic camera
D	disparity
d_c	distance from the lens array to the object plane in the integral imaging capture
d_d	display-to-combiner distance
d_m	combiner-to-eye distance
d_r	distance from the lens array to the reconstruction plane in the integral imaging reconstruction
d_v	viewing distance
δ_y	subpixel pitch of the LCD
$F.$	Fourier transform
f	focal length
$I(.)$	illumination function
k	the number of nearest microlenses to be integrated

$L(.)$	light field function
l_1	distance from the sensor to microlenses in the defocused plenoptic camera
l_2	distance from the microlenses to the main lens in the defocused plenoptic camera
l_c	distance from the sensor to the lens array in the integral imaging capture
l_r	distance from the display to the lens array in the integral imaging reconstruction
λ	wavelength
N_{px}, N_{py}	perceived resolution
N_v	number of views
Ω_y, Ω_t	Fourier domain axes
s	the slope of the shear
T_M	main lens pitch
T_m	microlens pitch
(θ, ϕ)	incident angle of a light ray
W_e	eye pupil size
W_v	eyebox
X_t	lens pitch
X_t'	multiplexing period
X_v	viewing pitch
X_y	pixel size
X_y'	magnified pixel size
z_0	recentering distance
z_m	display DoF limit
z_l	lens thickness
z_{min}, z_{max}	scene depth boundaries
z_s	smooth motion parallax limit

1. INTRODUCTION

From the ancient Greek till today, the philosophical discussion about reality has been always active in one or another form. People have been trying to understand the surrounding world and simulate (visually) the nature. Due to the technological developments over the last centuries, we are much closer than ever to the ability to recreate the reality by visual means. An important milestone in this respect, is the conventional photography, invented in the mid-1820s. While it brought memories into visual artefacts, it was still far from realistic, as it represented the scenes only in 2D. Interestingly enough, the inventions of the first 3D displays came around the same time. Studies on binocular vision, i.e. the visual perception by two eyes, can be traced back to works by Euclid and Leonardo Da Vinci [64], however, it was Sir Charles Wheatstone who coined the term stereopsis and designed the first stereoscope in 1838 [66]. Later on, in 1856, Sir David Brewster introduced the lens based two-view display, which he defined as lenticular stereoscope [9]. Gabriel Lippmann carried the stereo vision one step further by suggesting integral photography in 1908 [42]. Instead of only left and right views, he proposed to captured the scene by many small cameras from various points of views. Since then, 3D photography and 3D displays have been active research areas.

Answering the question about how humans perceive and understand the surrounding world goes through answering the question how the human visual system perceives light. Light is the medium carrying information about the world around us and as such it is of key importance in defining the elements of vision. Adelson and Bergen [1] discussed the elements of early vision and structured the visual information by defining single function, which they called plenoptic function. The plenoptic function describes the intensity of all rays propagating through all points in all directions. Thus, it defines a ray using 7 variables: location in 3D space, direction in 2D (spatial) angles, time and wavelength. Under certain assumptions, discussed in Chapter 2, the 7D plenoptic function can be reduced to 4D light field (LF) representation [21, 39]. It should be noted that this description models light as a collection of rays adopting the ray optics formalism. For more accurate representation of the nature of light, physical (wave) optics or electromagnetism should be applied. However, ray (geometrical) optics solutions provide sufficient accuracy for most of today's capture

and visualization techniques. In this thesis, we also utilize ray optics principles to model the behavior of light.

The plenoptic function is continuous due to the continuous nature of the light. However, it is impractical to describe the scene from every possible point [1]. Instead, by utilizing various capture systems, the plenoptic function is sampled and then the LF analysis is carried out based on these samples. Such capture systems are classified in [38] as integral photography [42], moving single camera [39], camera arrays [67], plenoptic cameras [2, 51, 44], light field microscopy [40], etc. In order to comprehend the LF data captured by each system, sampling process should be investigated elaborately. Therefore, as a preliminary study, fundamentals of the plenoptic sampling based on the geometrical analysis are discussed in Chap. 2 of the thesis.

It is essential to enlighten the discretization concept of the most simple LF capture setup to understand how LF data is relevant to the 3D information and how to process it further. However, the sampling procedure of each capture technique mentioned above should be analyzed individually as well, since each of them samples the plenoptic data in a different way. In this thesis, the objects of interest amongst the capture techniques are limited to the plenoptic cameras. Such devices are beneficial, since a plenoptic camera is capable of capturing the LF information in single shot, after which various post-processing algorithms can be applied, such as novel view synthesis, refocusing, noise reduction, depth estimation, etc. In the presented work, we are particularly interested in simulating the defocused plenoptic camera, in order to obtain ground-truth LF data for further processing. As a result, two different types of the plenoptic cameras are reviewed briefly under the discussions of the capture methods in Chap. 3, and a simulation algorithm implemented for the defocused plenoptic camera design is introduced along with the experimental results.

Plenoptic reconstruction and 3D scene display is as popular problem as the LF capture techniques. The conventional (2D) displays provide the texture information of a scene to the viewer; however, they are incapable of bringing the depth information into visualization, which is essential for more realistic experience. To address this issue, various approaches have been suggested for decades. In-depth review of such techniques are available in the literature [26, 17, 25]. In [25], the parallax characteristics of the displays are chosen to be the main component of the classification. The taxonomy of such techniques in their paper consists of the two-view, horizontal parallax and full parallax displays. Two-views include the eyeglasses based displays, stereoscopes and autostereoscopy. Horizontal parallax imaging systems are also named as the autostereoscopic displays, which can be multiview (MV) [16], or

super multiview (SMV) [33, 61]. Full parallax devices are given in three categories, namely integral imaging (InI) [69, 55], holographic displays [37], and volumetric displays [6]. In [17], on the other hand, two categories are considered as the main division: eyeglass-based binocular stereo displays and autostereoscopic 3D imaging techniques that do not require any special gear. Autostereoscopic displays are then divided into the multiview, volumetric and digital holographic displays. InI displays are reviewed under the refraction based multiview displays. Among all the display techniques mentioned above, we examine the refraction based multiview displays. We particularly discuss SMV and InI systems that utilize lens arrays to achieve 3D display capability. Since both methods are capable of providing the necessary depth cues in ray optics formalism [23, 47], they provide attractive solutions in various 3D display applications, e.g. virtual or augmented reality. Moreover, as it will be discussed later in Chap. 4, there exists a correspondence between the sampling structure of the plenoptic cameras and optical reconstruction of the reviewed displays, which constructs a link between Chap. 3 and Chap. 4.

As a summary, the construction of this thesis is as follows: In Chap. 2, the details of the plenoptic sampling and the discussions related to the minimum sampling rate are presented. Chap. 3 is devoted to the capture systems, where the plenoptic cameras are reviewed and the defocused plenoptic camera simulation algorithm is evaluated. After that, Chap. 4 presents the analysis of the SMV and the InI displays, as well as the examination of the developed prototype SMV display. Finally, the conclusion and the future work are presented in Chap. 5.

2. PLENOPTIC SAMPLING

The 7D plenoptic function $L(x, y, z, \theta, \phi, \lambda, t)$ models the intensity distribution of light in 3D space using collection of light rays, where at a given time t, each ray of wavelength λ is parametrized by a starting point (x, y, z) and direction (θ, ϕ) of propagation [1]. Under the assumptions of a static scene and fixed wavelength, plenoptic function can be reduced to 5D LF function. Restricting the set of rays to the ones that are propagating to e.g. $+z$ direction, it can be further reduced to the 4D LF, which is described by either the two-plane parametrization or space-angle representation [39]. Throughout the thesis, the two-plane parametrization is used. The two plane LF parametrization, also defined as light slab by [39], is illustrated in Fig. 2.1. By defining the intersection points of the light ray with both planes, both spatial and angular information of the LF are coded.

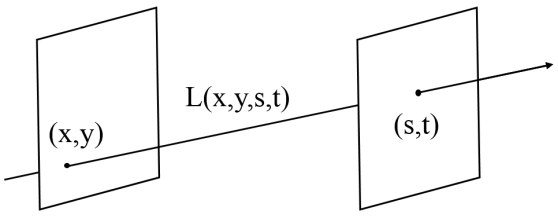


Figure 2.1 *Two plane parametrization of the light field.*

The continuous nature of the light brings continuous LF function as parametrized above into consideration. However, the explicit geometric information of a given scene can be reconstructed from discretely sampled LF data in practice. In order to further analyze the required sampling rate for accurate reconstruction, it is beneficial to introduce the discrete LF capture process first.

The most simple LF capture setup, in which an array of pinhole cameras is used, is illustrated in Fig. 2.2, where f is the focal length of the pinhole cameras, t is the camera plane, y is the sensor plane, and z is the distance to a plane in the scene. For the sake of simplicity, the 2D slice (y, t) of the 4D light slab (x, y, s, t) will be used throughout the thesis. Generalization to 4D is straightforward. It is also worth to note at this point that although the two-plane parametrization is employed in this thesis, the terms angular and spatial resolution are utilized as well to describe

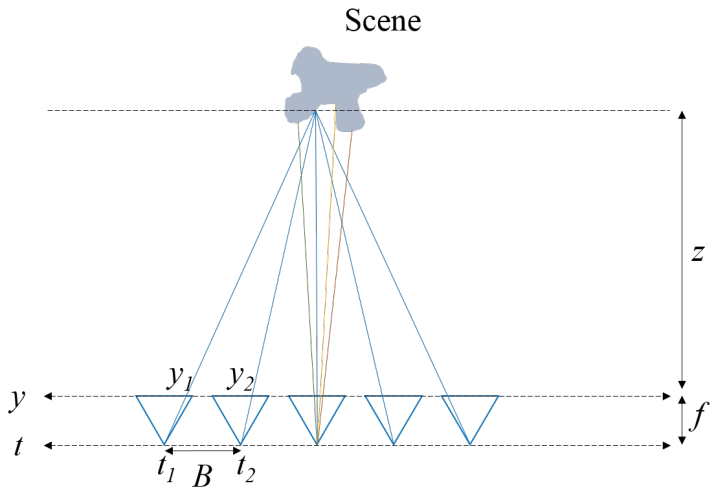


Figure 2.2 *Typical light field capture setup with an array of cameras.*

the capture or display processes. Therefore, before proceeding any further, the link between the space-angle and the two-plane parametrization should be constructed. Consider the object point in Fig. 2.2, from which the rays of different directions are captured by each camera. Since the rays are propagating from the same spatial location on the scene with varying angles to the cameras, the sampling rate of the camera array on $t - axis$ defines the angular sampling rate of the captured scene. Furthermore, the sampling rate on $y - axis$ for each camera corresponds to the texture information of the scene, which can be expressed as the spatial sampling of the scene (e.g. rays of different colors in the central camera capture discrete locations on the scene). Please note that each camera image (i.e. spatial information from a fixed angle) is referred as to perspective view, or parallax image, interchangeably throughout the thesis.

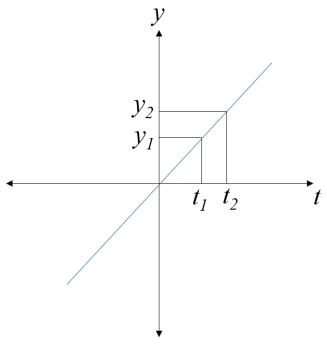


Figure 2.3 *Two-plane parametrization of the continuous LF of a constant depth plane.*

In Fig. 2.2, the centers of projections of cameras are placed on $t - axis$ and the sensor planes are represented by $y - axis$. Let us define the sensor coordinates on

$y-axis$ with respect to the camera coordinates on $t-axis$, i.e. each pixel inside the sensor of the camera at t_i has the $y-coordinate$ relative to t_i. Assume that we have continuous set of cameras, i.e. continuous LF is captured. If we stack the corresponding pixels of each camera that captures the same object point at the constant depth plane, we will obtain a line, as illustrated in Fig. 2.3. Please note that we assume a Lambertian scene for simplicity, in which case the resulting line is monochromatic. Each such line is called as epipolar line, where collection of all such lines for different scene points at different locations form the so-called epipolar image (EPI) [7]. The equation of the line can be defined noting the similar triangles in Fig. 2.2, such that

$$y - y_1 = (t - t_1)\frac{f}{z}. \tag{2.1}$$

If we define $D = y_2 - y_1$ as disparity and $B = t_2 - t_1$ as the baseline between adjacent cameras, then we can conclude that $D = Bf/z$. We can see that the disparity D is inversely proportional with the object distance. Thus, in a sense, the depth information is coded in correspondences between different camera images through disparity values.

The spectral support of the continuous LF of a scene at constant depth is given as in Fig. 2.4(a) [12]. The equation of the constant-depth line in the figure, $f/z\Omega_y + \Omega_t = 0$, indicates that the slope of the line is dependent on the depth of the plane. In other words, different planes at different depths correspond different lines. The Fourier domain representation of a scene with multiple depth planes, therefore, is shown in Fig. 2.4(b). Then, it can be concluded that for a given scene with finite depth boundaries, the continuous LF is bandlimited in the spectral domain. The boundaries of the spectral support are the lines that correspond to the maximum and minimum depths. However, please note that this condition is valid for occlusion-free scenes. In the case of occlusion, the same object point can not be seen from all views, which creates discontinuity in the epipolar line. The effect of occlusion in LF sampling is addressed in several studies [41, 71]. Here in this discussion, it is assumed that there is no occlusion in the scene. Fig. 2.4(c) illustrates LF discretization in the Fourier domain for a discrete set of cameras and sensor pixels, where the spatial and angular sampling rates are dependent on the sensor pixel pitch X_y and the baseline B, respectively. The rectangular box shows the band of the filter that enable reconstruction of the continuous LF via conventional bandlimited reconstruction, where X_y and B correspond to sampling at Nyquist rate.

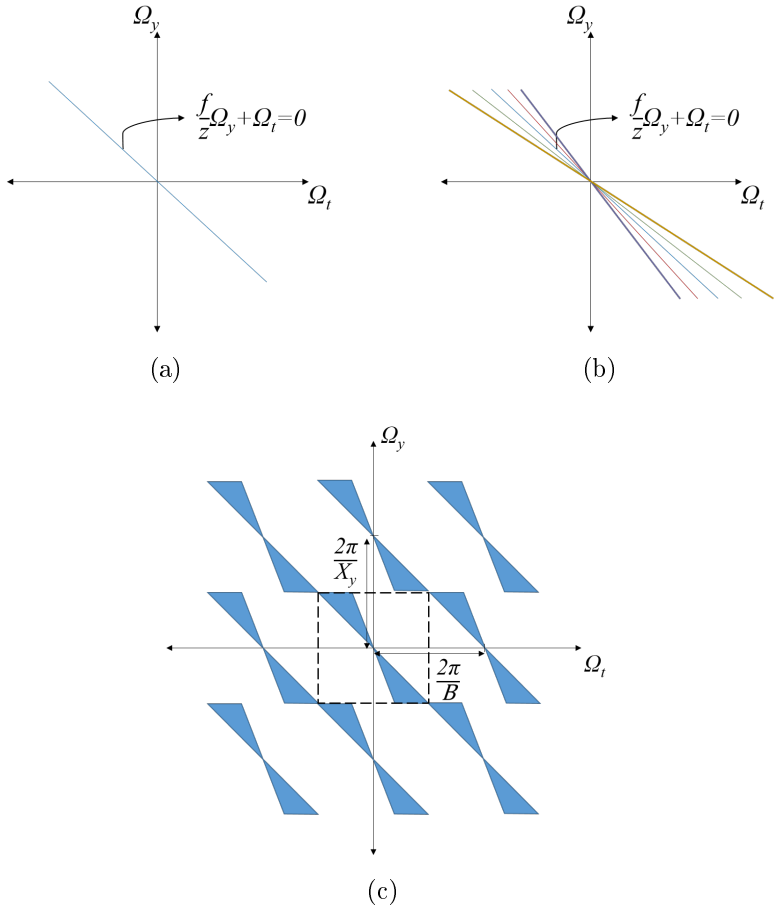


Figure 2.4 *Fourier domain analysis of the LF with constant depth (a), multiple depth planes (b), and discrete LF with multiple depth planes (c).*

2.1 Recentered Camera Array

The capture process illustrated in Fig. 2.2 consists of the cameras of which the center of projections and the center of the sensors are aligned. It should be noted that in this case, the field of views (FOVs) of cameras intersect in a limited area, which wastes notable amount of pixels especially for the closer scenes (e.g. macrophotography). Several techniques might be applied to increase the FOV intersection, amongst which is recentering the camera sensors onto a plane within the scene boundaries. Fig. 2.5 illustrates the LF capture using cameras with the recentered sensors. Let us consider the camera placed at t_n in the figure. If the sensor is shifted by $Y_N = t_n f/z_0$, the center of the captured image then corresponds to the center of the recentered plane v. It is illustrated that the same index pixel of each camera captures exact same locations on v plane, which means there is no disparity between the cameras on that plane. The disparity is positive for the planes closer and negative for the planes further to the cameras compared to v. Please note that with the

recentering camera array configuration, redundant disparity values are omitted. For a scene with the disparity range of [50, 55] pixels, for instance, if the cameras are recentered with respect to the plane exhibiting 52 pixels disparity, the new range then becomes [-2, 3] pixels, decreasing the waste of the pixels significantly. Please note that recentering the cameras actually corresponds to shearing the sampling band in the angular axis of the Fourier domain as shown in Fig. 2.6(a). As a result, as illustrated in Fig. 2.6(b) and Fig. 2.6(c), one can relive the sampling requirement. In particular, the angular sampling rate determined by the baseline can be reduced, i.e. the same LF content can be captured by a sparser set of cameras with baseline $B' > B$.

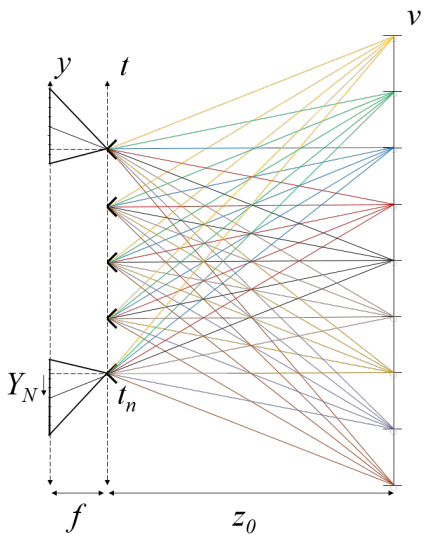


Figure 2.5 *Recentering the camera sensors on a plane in the scene in order to eliminate the redundant disparity.*

We conceptually demonstrated the LF sampling in the case of conventional and recentering cameras. In the following section, a more structured analysis related to the minimum (angular) sampling rate is presented.

2.2 Densely Sampled Light Field

There has been several studies investigating the minimum camera spacing utilizing scene geometry [41], or ray-space signal analysis [12, 28, 71] in details. In the following, the condition satisfying the densely sampled light field (DSLF), which is explained in these studies, is given under the assumptions of Lambertian, occlusion-free scene with finite depth boundaries.

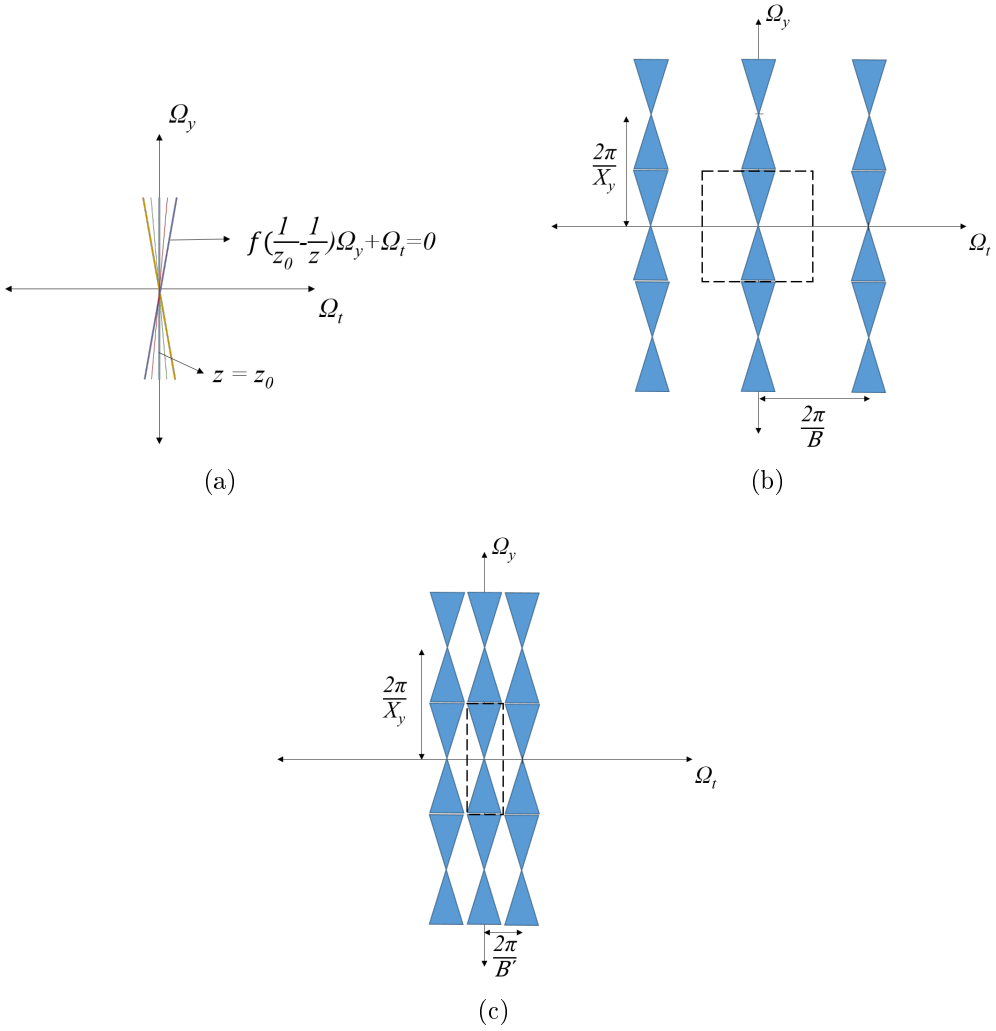


Figure 2.6 *Fourier domain analysis of the LF captured by the recentered camera array. The continuous LF (a) is captured with the baseline values of B (b) and B'>B (c).*

Assume a given scene with fixed depth boundary values $[z_{min}, z_{max}]$, where $z = 0$ represents the camera plane (t), and each camera is recentered with respect to the plane at $z = z_0$. The DSLF capture of such scene is guaranteed if the disparity value between the adjacent cameras are bounded to [-1,1] pixels [12], after which the continuous LF can be reconstructed from discrete samples using bilinear interpolation. In order to provide the DSLF capture, i.e. to ensure that the maximum disparity value is one pixel within the scene boundaries given the camera parameters, the maximum baseline value between adjacent cameras, B, should be [12]

$$B = \min\left\{\frac{z_0 X_y z_{max}}{f(z_{max} - z_0)}, \frac{z_0 X_y z_{min}}{f(z_0 - z_{min})}\right\}, \tag{2.2}$$

where f is the focal length and X_y is the pixel size of the pinhole cameras. Choosing

z_0 outside of the depth range $[z_{min}, z_{max}]$ results in negative value in one of the arguments in Eq. 2.2, in which case the baseline B should be equal to the other (positive) argument. In order to keep B as large as possible, i.e. to capture the scene with the sparsest set of cameras, the recentering depth z_0 should be chosen as

$$z_0 = \frac{2z_{min}z_{max}}{z_{min} + z_{max}}, \tag{2.3}$$

which is also called as the optimal constant depth [41]. Please note that such choice of z_0 equalizes the disparity between the depths $z_{min} - z_0$ and $z_0 - z_{max}$. Using Eq. 2.3 to replace z_0 in Eq. 2.2, then, it can be observed that both arguments are equal to each other, which is

$$B = \frac{2X_y z_{min}z_{max}}{f(z_{max} - z_{min})}. \tag{2.4}$$

The significance of the DSLF is that it provides a structural framework to analyze the characteristics of the given scene. After reconstructing the continuous LF from regularly sampled rays, any desired ray can be resampled using linear interpolation, which then eases the implementation of various algorithms such as view synthesis, refocusing, super-resolution, etc. In the presented work, the sampling and the reconstruction framework based on the DSLF is utilized as well. In particular, during the implementation of the defocused plenoptic camera simulation algorithm, we rely on the captured DSLF in order to simulate the ray integration within the sensor pixels.

3. LENS ARRAY BASED 3D CAPTURE

In this chapter we focus on lens array based capture systems. In particular, two different (defocused and focused) plenoptic camera configurations are reviewed. The sampling process and the spatio-angular resolution trade-offs are evaluated and possible application areas are discussed for both designs. Furthermore, a simulation algorithm is developed for the defocused plenoptic camera and the corresponding experimental results are given in details.

3.1 Plenoptic Cameras

3.1.1 Defocused Plenoptic Camera

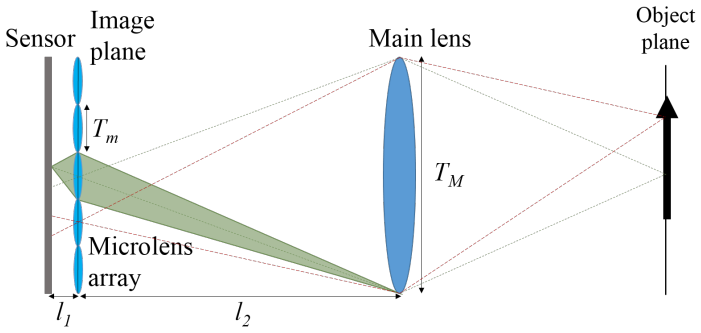


Figure 3.1 *Defocused plenoptic camera design. Main lens focuses the scene on the microlens plane and microlenses are focused on the main lens (adapted from [3]).*

Traditional plenoptic camera design, the so-called defocused plenoptic camera, was first introduced by Adelson and Wang in 1992 [2]. After that, Ng et al. improved this design to produce a portable plenoptic camera [51]. Such camera has one main lens and one microlens array on the image plane of the main lens, as illustrated in Fig. 3.1. In conventional 2D cameras, the sensor plane is placed on the image plane; however, in (defocused) plenoptic camera design, the sensor plane is placed behind the microlens array such that each microlens focuses on the main lens plane, or practically infinity. This design is biologically described in [51] as if an insect eye

is placed on the retinal plane of human eye. By this way, the angular LF information is obtained as well, which is lost in conventional 2D photography.

Traditional 2D cameras provide high spatial resolution; however, single 2D image has a fixed focused plane and limited depth of field (DoF). The aperture size is inversely proportional to the image DoF, meaning that it can be extended by choosing smaller aperture. With small aperture, though, the exposure time should be increased in order to get enough amount of light, in which case the motion blur should be taken into account. These problems constitute example motivations for the plenoptic cameras [50]. As it is briefly explained in the following, with the LF data captured by single shot of the plenoptic camera it is possible to extend the image DoF or refocus the image on another plane, as well as to synthesize novel views within the main lens aperture. In this section, the sampling procedure is investigated and the resolution trade-off is also discussed.

The defocused plenoptic camera layout is shown in Fig. 3.1. The dashed lines in the figure indicate that the elemental images, i.e. small images behind each microlens, consist of the pixels which carry the angular information propagating from the same spatial location on the object plane of the main lens, while the microlenses obtain the information of the spatial distribution. It should also be noted that the spread of each elemental image depends upon the aperture size of the main lens [51]. If the main lens aperture is large, the elemental images covers wider area in the sensor, while smaller aperture results in smaller elemental images, from the similar triangles in the figure. Since there is no physical barrier between the microlenses, if the main lens aperture is large so that the elemental image size is larger than the microlens aperture, the adjacent images overlap, which results in cross-talk between them. If the main lens aperture is too small, on the other hand, the pixels on the border will not capture the LF information. Thus, LF camera will not be utilized efficiently. As a result, in order to employ the pixels under the microlenses most effectively, the (image-side) $f-$number of the main lens should be matched with the (image-side) $f-$number of the microlenses, such that [51]

$$T_m/l_1 = T_M/l_2, \tag{3.1}$$

where l_1 is the distance between the sensor and the microlens array plane (image distance of microlenses), l_2 is the distance between the microlens array plane and the main lens (main lens image distance), T_m is the microlens pitch, and T_M is the main lens pitch.

The significance of subperture image analysis is discussed in [2, 51]. As they demon-

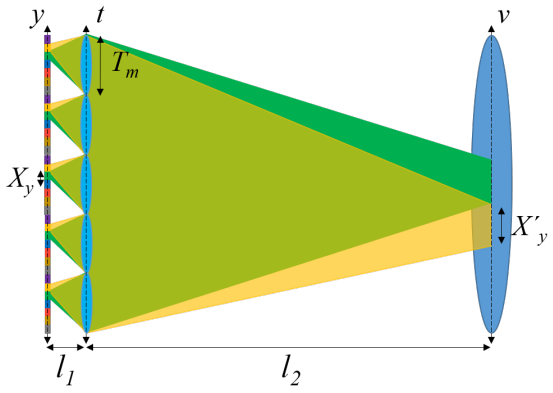


Figure 3.2 *Image formation and perspective images in defocused plenoptic camera. Each pixel in elemental image captures the corresponding pixel of one perspective image.*

strated, subaperture images are formed by taking the corresponding pixels under each microlens. Fig. 3.2 shows two adjacent images with different colors. We can conclude from the figure that each subaperture image corresponds to a perspective projection of which the center of projection is located at the main lens plane. Since in the ideal case the microlenses are focused on the principal plane of the main lens [51], each pixel maps the area $X'_y = X_y l_2/l_1$ within the main lens. Therefore, the aperture size of one perspective image, which is also the smallest subaperture size, is equal to X'_y. The number of the perspective images are equal to the number of pixels inside one elemental image, each capturing the scene from different angle. This configuration in a way fixes the resolution constraint; the spatial resolution of each perspective image is as much as the number of microlenses.

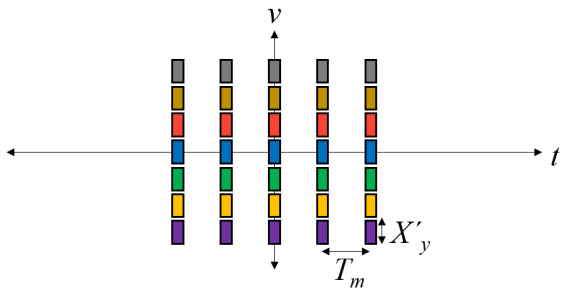


Figure 3.3 *Sampling in defocused plenoptic camera, where v is the main lens plane and t is the microlens array plane.*

Having the parametrization in Fig. 3.2, the defocused plenoptic camera samples the LF as illustrated in Fig. 3.3. The v and t axes represent the spatial locations on the main lens plane and the microlens array plane, respectively. Please note that the sampling grid does not represent the integration fully. Since each pixel integrates all

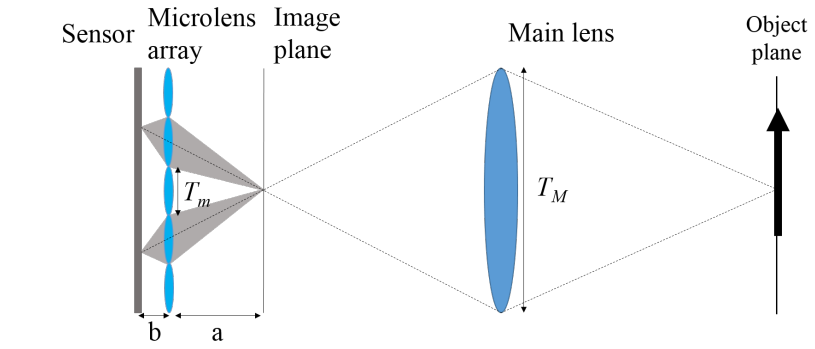
the rays along the microlens aperture in reality, the width of each sample in $t-axis$ is T_m, leaving no gap between the samples. Fig. 3.3 is drawn with gaps for clarity.

Each color represents one perspective image, which has its center of projection on the main lens plane, while the sensor is on the microlens array plane. In order to construct a 2D image with a specified subaperture size and position, the corresponding samples along the $v-axis$ (i.e. corresponding colors) should be integrated. Since the samples are aligned on each axis, it is straightforward to observe that the resolution of the integrated subaperture image is equal to the resolution of one perspective image, which is the number of microlenses. Though the subaperture image resolution is restricted by the microlenses, it can be improved by super resolution techniques. In fact, this problem can be considered as high resolution image rendering from a set of low resolution (sub aperture) images. Several approaches have been proposed for this problem in the literature [5, 58, 62].

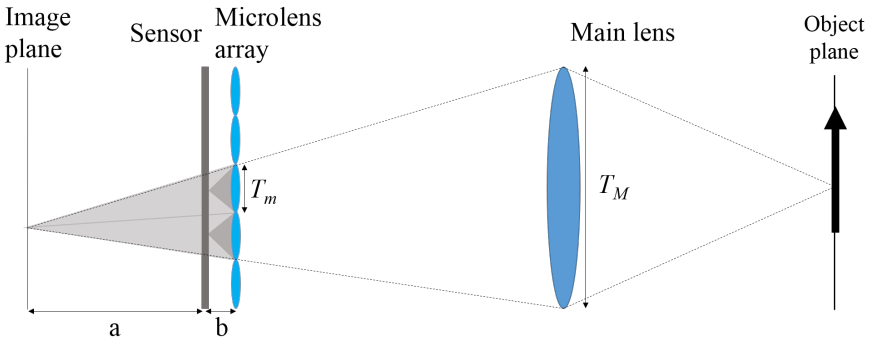
Fig. 3.3 gives the sampling on $v-t$ coordinates. There, reconstructed 2D image by integrating the perspective images along $v-axis$ is equivalent to a conventional 2D image of which the lens is positioned at v plane and the sensor is positioned at t plane. The corresponding focused plane can then be found with the lens equation. It is also important to note that, by simply recalculating the rays with $v'-t'$ parametrization on different synthetic lens and sensor planes, LF can be post-processed and synthetic images can be generated [51]. For instance, changing t' results in digital refocusing. Shifting t' in fact corresponds to shifting and adding the subaperture images. The examples to such applications will be further discussed in Section 3.2. More detailed information and formulation of the synthetic aperture photography can be found in [51].

3.1.2 Focused Plenoptic Camera

Although the defocused plenoptic cameras are attractive due to their solution to the imaging problems of the conventional 2D cameras by capturing the LF, as a trade-off they suffer from low spatial resolution. Focused plenoptic camera has been mainly proposed to address this spatial resolution issue and aimed to capture denser spatial information at the cost of resolution in angular information [44, 20]. As it can be seen in Fig. 3.4, in the focused plenoptic camera design, instead of being focused on the main lens (i.e. optical infinity), microlenses are focused on the image plane of the main lens. Two different focusing is achievable either by placing the image plane in front of the microlens array as in Fig. 3.4(a), which is similar to the Keplerian telescope design, or creating virtual elemental images by placing the



(a) Keplerian design.



(b) Galilean design.

Figure 3.4 *Two different focused plenoptic camera designs. Microlenses are focused on the image plane of the main lens, which can be in front of (a) or behind (b) the microlens array plane (adapted from [3]).*

image plane behind the microlens array as in Fig. 3.4(b), which is similar to the Galilean telescope configuration.

Considering the illustrations of the focused plenoptic cameras in Fig. 3.4, it can be concluded that the LF sampling process of such devices is different than that of the defocused plenoptic camera. In the defocused plenoptic camera, the LF data can be defined as the collection of the perspective images of the scene seen from the main lens plane. Any 2D image with a specified aperture size can then be constructed by integrating these perspective images, as discussed in details in Section 3.1.1. In the focused plenoptic camera, however, the object is mapped by the main lens in front of (Keplerian, real image), or behind (Galilean, virtual image) the microlenses, from where the 2D perspective micro-images are stored by each microlens [44]. Therefore, the spatial information along the object plane in Fig. 3.4 is first transferred to the image plane, then obtained by the pixels inside each elemental image. On the other hand, the angular information is captured by the perspective images along the microlens array plane. This difference in sampling changes the final spatial

resolution of the 2D image rendered from the captured LF. In order to render a high resolution 2D image of the scene by the main lens of the focused plenoptic camera, the low resolution elemental (perspective) images are blended [44, 20]. In the following, the high resolution 2D image formation is discussed.

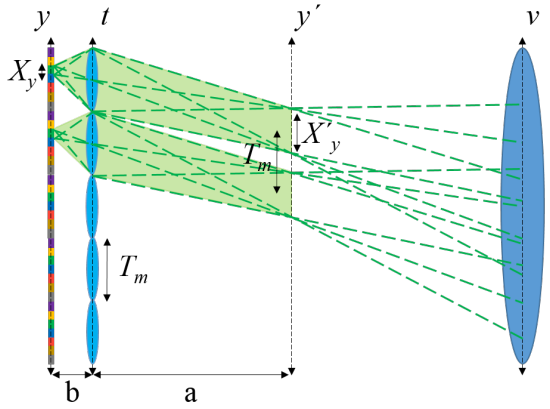


Figure 3.5 *LF capture setup in the focused plenoptic camera. Each elemental image captures the scene (transformed by the main lens) at different angle.*

The parameters and the LF sampling structure of the Keplerian-design focused plenoptic camera is illustrated in Fig. 3.5. The distance between the microlenses and the sensor planes is denoted by b. The microlenses are focused at y' at distance a from the microlens plane, which can be calculated from the lens equation $1/b+1/a = 1/f_m$, where f_m is the focal length of microlenses. Let us choose t and y' planes for the two-plane LF parametrization. Each pixel inside an elemental image captures the ray bundle propagating from the area represented by X'_y at y', which is the magnification of the pixel size, $X'_y = X_y b/a$, through the microlens aperture T_m at t, as illustrated with the shaded regions. The distance between the central rays of adjacent elemental images is T_m as well. Therefore, assigning different colors to each pixel inside one elemental image, the resulting sampling grid in $t-y'$ representation is given in Fig. 3.6. Please note that the figure is again drawn with the gaps between the samples for clarity.

As it was mentioned earlier and can be seen in Fig. 3.4, the spatial coordinates of the light rays propagating from the object plane are mapped to the the image plane y' to be captured by the microlenses. Thus, integrating the samples along the $t-axis$, which represents the angular distribution of rays, creates the 2D image of the scene focused at the object plane. The width of each sample is X'_y on $y'-axis$, which defines the pixel size of the rendered 2D image. As it was given earlier, $X'_y = X_y a/b$; therefore, the final spatial resolution is achieved to be b/a times the sensor resolution [44]. Please note that this resolution is obtained with basic rendering algorithm

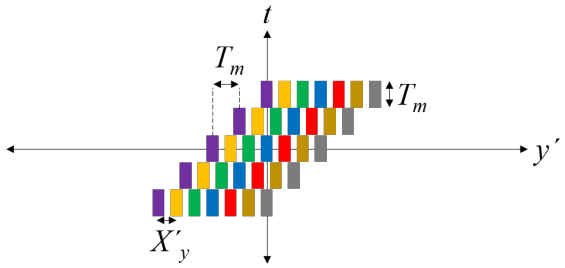


Figure 3.6 *Sampling in focused plenoptic camera, where t is the microlens array plane and y´ is the image plane of the main lens (also the object plane of the microlenses).*

based on the magnification of the elemental image resolution. As it can be clearly seen from Fig. 3.6, the samples are not aligned vertically. Therefore, a denser spatial resolution can be achieved by using the subpixel disparity between adjacent views and defining a finer grid on y' [18]. The reader is referred to [44, 20, 18, 19] for more detailed analysis on image rendering and the super resolution algorithms in focused plenoptic cameras.

Due to the presented differences in the designs, the defocused and the focused plenoptic cameras have different pros and cons, which should be considered before using them for the intended application. In few of the post-processing applications, such as refocusing or viewpoint change, denser angular sampling provides better results, in which case the defocused plenoptic camera is more suitable. In some other implementations, e.g. depth estimation, the algorithm can be enhanced with better spatial resolution, for which the focused plenoptic camera can be considered.

3.2 Defocused Plenoptic Camera Simulation Utilizing DSLF

In order to utilize the raw LF data for further applications such as refocusing, precalibration of the physical camera is needed. The calibration algorithms already exist in the literature [14, 31]. However, the accurate calibration is usually a hard task, which requires extra effort. Thus, it complicates the utilization of the camera. In order to examine post-processing applications in a reliable way, a simulation algorithm that provides the ground-truth plenoptic data for an animated scene is valuable.

In the literature, various plenoptic camera simulation algorithms exist [57, 72, 43]. Since each pixel on the sensor integrates the ray bundle propagating through the microlens aperture to the pixels, the accurate implementation requires simulation of these integrations on both planes. The existing algorithms mainly aim to simulate

these integrations by forming randomly oversampled grids within the microlens and the pixel boundaries. Then, by superposing the intensities of rays defined between sample points, the integration is approximated. In [57], backward ray tracing, i.e. tracing a ray from sensor plane to the object points in the scene, is used to calculate the intensity of each ray. The process includes the camera lens system simulation and the path tracing algorithm, from camera to the scene. Apart from their implementation, there exists simulation algorithms utilizing the forward ray tracing, i.e. from scene to the sensor plane, such as in [72, 43]. In [43], Liu et. al. simulate the plenoptic camera based on ray splitting utilizing Monte Carlo method, where the plenoptic imaging is simulated successfully and consequences of the design error can be analyzed by comparison of the physical camera to their algorithm. Here in this thesis we rather aim to simulate the defocused plenoptic camera by utilizing the 2D pinhole renderings of a simulated scene in a computer graphics rendering software, e.g. Blender, based on our work in [3]. As discussed in Chapter 2, the DSLF can reconstruct the continuous LF of a given Lambertian and occlusion-free scene by linear interpolation of the discrete samples. Since each sensor pixel integrates the rays within its boundaries, if linearly interpolated continuous LF is reconstructed, the explicit integration for each pixel can then be derived. The integration gives more exact solution compared to the superposition of the samples in an overly sampled sensor pixel. In the following, implementation details are presented, as well as the experimental results and the future work.

3.2.1 DSLF Capture Setup

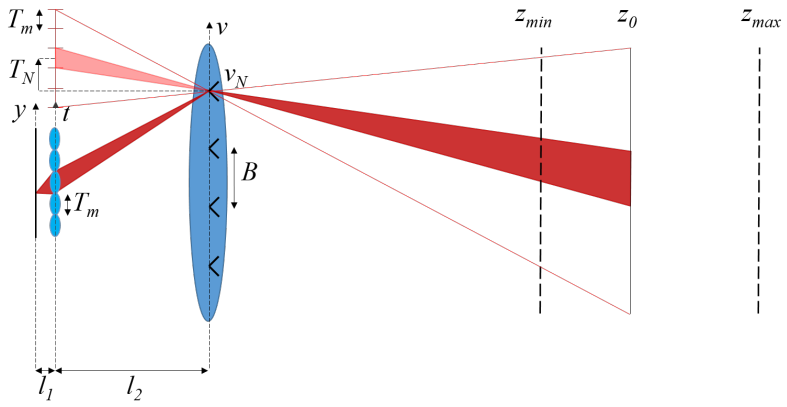


Figure 3.7 *Dense light field capture process for defocused plenoptic camera simulation (from [3]).*

The configuration related to the first part of the algorithm, namely DSLF capture inside the main lens aperture, is explained in Fig. 3.7. The center of projection of

the pinhole cameras are placed on the main lens plane of the defocused plenoptic camera, while the sensors are placed on the microlens array plane. The pinhole cameras are then recentered with respect to the object plane of the main lens z_0, which can be calculated by the lens equation using the main lens focal length f_M, and the distance between microlenses and main lens l_2 (i.e. image distance), such that $z_0 = (1/f_M - 1/l_2)^{-1}$. The sensor of each pinhole image should then be shifted according to recentering equation, i.e. for a pinhole camera of which the center of projection is at v_N, the sensor shift should be $T_N = v_N l_2/z_0$. Finally, by setting the pixel size as the microlens pitch T_m, DSLF capture setup is established. Please note that the microlenses are assumed to be square in this setup. For a simulation with the microlenses of different shape, the pixel configuration in the DSLF capture setup should be changed accordingly. As it is illustrated in Fig. 3.7, the dark red ray bundle coming from the scene and illuminating the third microlens after getting refracted from the main lens is captured by third pixel of the pinhole camera placed at v_N as the shaded red. If the DSLF captured by the pinhole cameras are integrated along the main lens aperture, the setup correctly simulates a conventional 2D camera, of which the lens is the main lens of the defocused plenoptic camera and pixels are the microlenses. Since each perspective image in the LF representation of the plenoptic data is the subaperture image within the main lens, it is straightforfard to observe that the plenoptic data of the given scene can be implemented by the integration of the DSLF within the defined boundaries. In the following, the explicit integration is discussed in more details.

3.2.2 Ray Integration

The intensity value of one sampling point on the sensor grid of the defocused plenoptic camera is the integration of the ray bundle propagating from the corresponding point on the main lens plane to the parent microlens, as illustrated in Fig. 3.8. For a pixel covering the area X_y, the total intensity is the collection of all the ray bundles within the area X'_y on the main lens plane. Since DSLF capture provides a structured framework for the LF capture, it is possible to derive this integration explicitly. In this section, derivation of the integral upon the interpolated continuous LF is discussed.

Fig. 3.8 illustrates two ray bundles propagating through the microlens t_i and captured by the boundary points of the pixel y_j. Each bundle corresponds to the ith pixel of a pinhole camera located at the main lens plane. Since we capture the DSLF as explained in Section 3.2.1, we can then reconstruct the continuous LF for t_i via the linear interpolation of the ith pixels of each (captured) pinhole image.

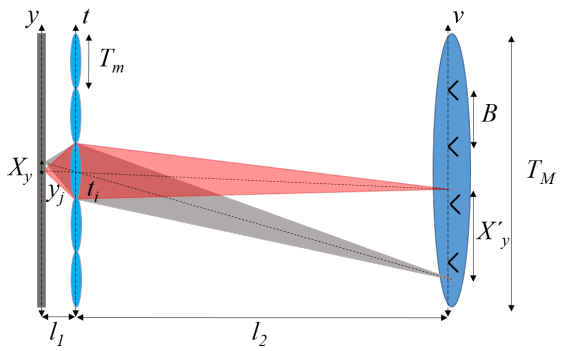


Figure 3.8 *Capture process for a pixel in the defocused plenoptic camera (from [3]).*

Such a continuous function is drawn in Fig. 3.9.

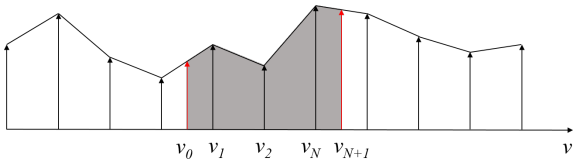


Figure 3.9 *Integration process for sensor grid; samples from pinhole images reconstructs the original signal, from which rays can be integrated within the pixel boundaries (from [3]).*

Assuming that the area X'_y in Fig. 3.8 is bounded by points v_0 and v_{N+1} on the main lens plane, the integration of the piecewise linear continuous LF gives the total intensity of the ray bundle propagating from t_i to y_j, $I(y_j, t_i)$ as [3]

$$I(y_j, t_i) = \int_{v_0}^{v_{N+1}} L(t_i, v)dv = \frac{1}{2}\sum_{n=0}^{N}(v_{n+1} - v_n)[L(t_i, v_n) + L(t_i, v_{n+1})], \tag{3.2}$$

where the continuous LF function $L(t_i, v)$ is interpolated by the corresponding pixels of the adjacent pinhole cameras $v_1, ..., v_N$. The maximum number of pinhole camera samples within X'_y can be derived as [3]

$$N = \left\lceil \frac{X_y l_2}{B l_1} \right\rceil. \tag{3.3}$$

If there is no physical barrier between the microlenses, the total intensity recorded by y_j, $I(y_j)$, is then the summation of $I(y_j, t_i)$ over k microlenses, such that

$$I(y_j) = \sum_{\mathrm{k}} I(y_j, t_i). \tag{3.4}$$

The maximum number of microlenses that can illuminate y_j, k, can be calculated

as

$$k = \left\lceil \frac{T_M}{T_m l_2 / l_1} \right\rceil. \tag{3.5}$$

Please note that the denominator in Eq. 3.5 refers to the $f-$number match condition in Eq. 3.1. If the $f-$number match is ensured, y_j captures the light rays propagating from only one microlens. However, if the main lens aperture is larger than the matched value, more than one microlens should be taken into account for each pixel in order to simulate the cross-talk accurately.

Although the formulations here are derived in 1D for the sake of simplicity, the 2D integral should be computed in order to find the total intensity value of a sensor pixel. Due to the separability of the 2D continuous LF, such integration can be performed separately on each axis.

The efficiency of the integration is that it eliminates the oversampling on the sensor plane. The existing algorithms mostly (randomly) oversamples both the sensor and microlens array grids to form the desired rays. While doing this, the definition of the oversampling factor is usually left ambiguous. Thus, in this sense, the integration utilizing DSLF provides more structured solution to the problem.

3.2.3 Experimental Results

During the implementation of the proposed algorithm, the process is divided into two main parts. In the first part, which is the DSLF capture, pre-designed scene in Blender, an open source rendering tool, is rendered by dense set of recentered pinhole cameras. In the second part, namely the integration process, the final pixel intensities are calculated. In order to compare the close-up photography to the conventional range, two scene designs with different minimum and maximum depths are examined. For the camera parameters, the design specification in [51, 3] are used for both scene. The sensor with 4144×4144 pixels resolution is placed $l_1 = 502.9\mu m$ behind 296×296 microlens array, with the focal length $f_m = 500\mu m$. The resolution of each elemental image is then 14×14 pixels. The pixel pitch of the sensor is $X_y = 9\mu m$, and the aperture size of one microlens is $T_m = 126\mu m$. The focal length of the main lens is $f_M = 80mm$.

For large-scale scene, which has the depth range of $z_{min} = 1m, z_{max} = 3m$ [3], the main lens of the camera is focused at $z_0 = 1.5m$. Therefore, the distance between the main lens plane and the microlens array plane is chosen to be $l_2 = (1/f_M - 1/z_0)^{-1} = 84.5mm$. The $f-$number match is ensured in the first scene. The resulting aperture of the main lens is $T_M = T_m l_2 / l_1 = 21,2mm$. In order to capture the DSLF, 15×15

pinhole cameras of 296×296 pixels resolution are recentered to z_0. The cameras are placed within the main lens aperture, making the baseline $B = 4.5mm$. It can be concluded by Eq. 2.2 that the capture setup provides the DSLF condition.

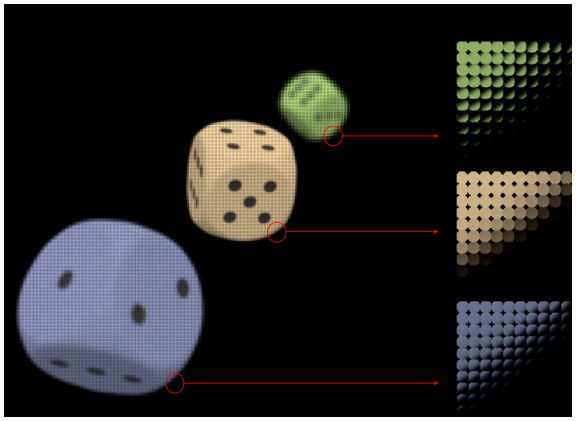


Figure 3.10 *Sensor image of the defocused plenoptic camera, where the scene depth range is 2-3m [3].*

The final sensor image as well as the zoomed in elemental images capturing three different depths are illustrated in Fig. 3.10. The circular shape of the main lens can be observed in the elemental images. Since the main lens is circular, the corner views are outside the main lens boundaries and the corner pixels of the elemental images are black, as illustrated in the figure. The total number of views captured within the main lens, which can also be defined as the angular resolution of the LF, is 10×10.

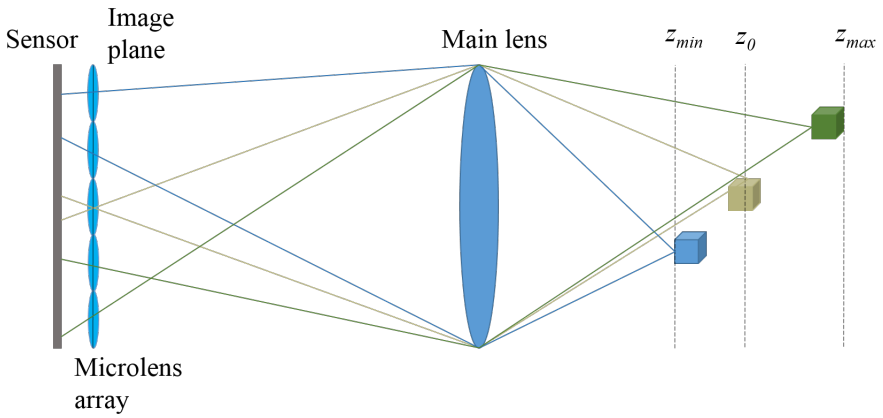
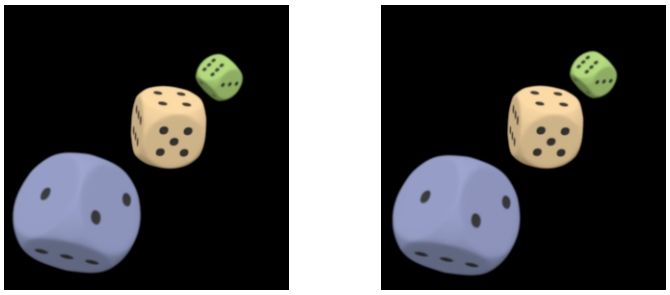


Figure 3.11 *Elemental image formation in the defocused plenoptic camera. The characteristics of the images are the opposite to each other when the closer and further objects are compared with respect to the recentering plane.*

The elemental image characteristics is observed to be varying with the depth of the objects, as consistent with the analysis in [2]. The intensity values of the pixels

capturing the beige dice, which is placed near the object plane of the main lens z_0 (recentering plane of the pinhole images as well), changing simultaneously. As the scene is Lambertian, the rays propagating from the same spatial point on the object plane with different angles are expected to have the same intensity values, which explains the condition of these elemental images. On the other hand, for the blue dice, which is closer to the plenoptic camera, the change in the adjacent elemental images are in the same direction with the gradient of the object. For the further object, green dice, the images are upside down, which results in opposite change with respect to the blue dice. In Fig. 3.11, this phenomenon can be seen clearly. Since the further object is focused in front of the microlenses by the main lens, the rays are crossed and propagate to the microlenses on the opposite side. For the closer objects, however, the image plane of the main lens is behind the microlens, therefore the rays are directly mapped. The depth map of the scene can be extracted by examining the amount of the pixel shift between the viewpoints.

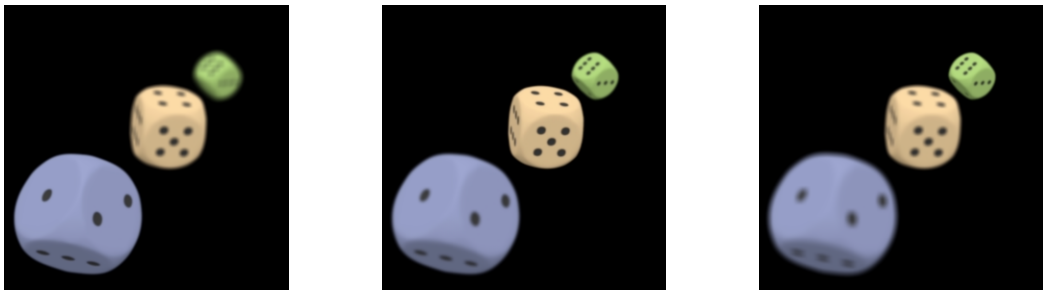


(a) (b)

Figure 3.12 *Perspective images of the first scene captured from the two ends of the main lens. The parallax is observed.*

After the formation of the sensor image, the 5D LF matrix in standard format is created, where first two are the angular dimensions (pixels behind each elemental image for the defocused plenoptic camera), third and fourth are the spatial dimensions (as much as the number of elemental images), and fifth dimension is the color information. The first two dimensions can also be defined as the positions of the perspective images on the main lens; any desired image can easily be obtained by extracting the correspondent columns and rows of the LF. Two such view images from the opposite sides of the main lens are shown as example in Fig. 3.12. The largest DoF achievable is the DoF of one perspective image. If the views are summed together, the 2D image of the scene within the main lens aperture is constructed, which has narrower DoF. Such image is sharp on the reconstruction plane and blurred in out-of-focus planes. Therefore, there is no parallax between view images the reconstruction plane, meaning that the perspective images are recentered there. Besides,

the parallax of the objects closer to the camera is opposite compared to the objects further than the reconstruction plane.



(a) (b) (c)

Figure 3.13 *Refocused images at z_{min} (a), z_0 (b), and z_{max} (c).*

The post-processing algorithms are also applied to the plenoptic image analyzed above, utilizing the light field toolbox [14]. The results of the digital refocusing algorithm are shown in Fig. 3.13. In particular, the planes where the dice are located are refocused, in which case the refocusing planes are around z_{min}, z_0, and z_{max}, respectively. The refocused images demonstrates the accuracy of the provided algorithm, as well as the validity of the simulated plenoptic image.

Figure 3.14 *Sensor image of the defocused plenoptic camera, where the scene depth range is 0.4-0.6m (Slava Z. ©2014 www.sketchfab.com, used under the Creative Commons Attribution license).*

In order to evaluate the simulation further, a second scene is designed with the depth range of $z_{min} = 0.4m, z_{max} = 0.6m$ [3]. The resulting sensor data and the zoomed in images are given in Fig. 3.14. The magnified elemental images capture the LF around z_{min}, z_0 and z_{max}, where z_0 is chosen to be $0.48m$ in accordance with Eq. 2.3. From the lens equation using f_M and z_0, l_2 is set as $96mm$. Please note that changing l_2 results in changing l_1 as well, since the microlenses are aimed to focus

on the main lens plane. Therefore, $l_1 = 502.6\mu m$ in the second configuration. The resulting (matched) aperture of the main lens is $T_M = 24,1mm$.

Since the scene is closer to the camera in the second design, denser array of pinhole cameras are needed for DSLF capture, according to Eq. 2.4. The baseline between the adjacent pinhole cameras are then $B = 3.2mm$, in which case an array of 21×21 pinhole cameras are needed within the main lens. The resolution of the pinhole images are again equal to the number of microlenses, which is 296×296.

In the discussions of DSLF, the scene is assumed to be Lambertian and occlusion-free. Although the first scene with dice satisfies these conditions, occlusion occurs in the second scene (e.g. around the mirror of the car), in which case the scene is not bandlimited, as discussed in Chapter 2. In theory, reconstruction of such scenes presents artifacts. However, as [41] discuss, with the minimum sampling rate for the occlusion-free scene, the artifacts that occur in the occluded scene are not noticeable in practice. Thus, they suggest that the sampling rate for a given scene can be decided in accordance with the depth range, regardless of the occlusions.

(a) (b)

Figure 3.15 *Perspective images of the second scene captured from the two ends of the main lens.*

The view images for the second scene are also shown in Fig. 3.15. The discussion of the perspective images for the first scene is valid for the second scene as well.

The results of the above-mentioned refocusing algorithm with the second scene design is seen in Fig. 3.16, with the refocusing distances are again z_{min}, z_0, and z_{max}. Please note that in the second scene, which is closer to the camera, the depth range is narrower ($20cm$) compared to the distant scene with the range of $2m$. It can then be concluded that for a closer scene, the post processing algorithms related to the motion parallax, such as refocusing, can be applied in a shallower depth range. In other words, the further the scene is, the larger depth range it should have in order to obtain the same digital refocusing results.

(a) (b) (c)

Figure 3.16 *Refocused images at z_{min} (a), z_0 (b), and z_{max} (c).*

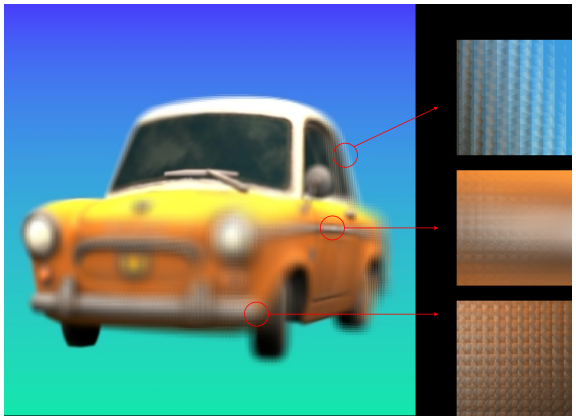


Figure 3.17 *Sensor image with plenoptic camera of which the main lens aperture is larger than the matched aperture value. The cross-talk between the elemental images are visible.*

We demonstrated the plenoptic images and the refocusing algorithms for two scenes with different depths. It was mentioned that for both of these demonstrations, the $f-$number match between the main lens and the microlenses is ensured. In order to further examine the simulation algorithm, the closer scene with the depth range of $z_{min} = 0.4m, z_{max} = 0.6m$ is captured with different camera parameters. In particular, the $f-$number of the main lens is chosen to be smaller than the $f-$number of the microlenses (i.e. the main lens aperture is larger). The resulting sensor image of the scene is shown in Fig. 3.17. The main lens aperture size is chosen as $T_M = 61.2mm$. In this case, the size of the elemental images are too large that they overlap, which creates cross-talk, as consistent with the discussions of Sec 3.1.1. The overlapping regions can be seen clearly in the zoomed in images.

The view images obtained from the plenoptic camera with large aperture are shown in Fig. 3.18. The ghost-image like artifacts can clearly be seen in the figure, decreasing the image quality. Please note that such artifacts are more clear at the scene boundaries. This phenomenon can be explained with the elemental image

(a) (b)

Figure 3.18 *Perspective images of the plenoptic data with cross-talk. Please note the aliasing effects.*

characteristics discussed earlier. Since the variety of information inside an elemental image increases when the object gets closer to boundaries, cross-talk is expected to create more artifacts in these regions.

(a) (b) (c)

Figure 3.19 *Refocused images at z_{min} (a), z_0 (b), and z_{max} (c) obtained using the LF data with cross-talk. The ghost images can be seen on the focused planes.*

The refocusing algorithm is applied to the LF data affected by cross-talk as well, and the results are illustrated in Fig. 3.19. The ghost images are obvious when the camera is focused on the front and the back sides of the car. These artifacts vanish when the object is out of focus, due to blurring.

4. LENS ARRAY BASED 3D DISPLAYS

In this chapter, InI and SMV displays are discussed in details as examples of autostereoscopic display techniques that utilize lens arrays. In addition to theoretical discussion, we present a prototype SMV head-up display system as practical demonstration of SMV display technique. InI and SMV provide attractive 3D display solutions with the ability of creating all physiological depth cues correctly, thus delivering realistic viewing experience. Such depth cues can be divided into four main categories, namely vergence, binocular disparity, accommodation and motion parallax [54]. Vergence is the movement of left and right eyes to each other to fixate on the object. Difference between the sensed images in both eyes, binocular disparity, then maintains the binocular vision. Accommodation, on the other hand, is a monocular cue, specifically effective for short distances ($1 \sim 3m$). It can be defined as the change in the optical power of the eye lens in order to change the focus distance. The last depth cue, motion parallax is created by relative motion of the objects at different depths, when the head is moved. It is effective in both short and long distance perception. Please note that the ability to provide smooth motion parallax and correct accommodation (focus) cues constitute the two critical aspects of 3D perception that are not able to be addressed by all 3D display techniques, e.g. stereoscopic displays or conventional multiview displays.

4.1 Integral Imaging

First suggested by Gabriel Lippmann as integral photography in 1908 [42], InI constitutes one of the oldest 3D display techniques. InI technique actually provide a solution for both 3D imaging and display problems by utilizing two dimensional microlens array in front of the sensor and a 2D display, respectively. The capture and display setup of InI is illustrated in Fig. 4.1.

In the capture stage, each microlens spatially samples the LF incident on the microlens array plane and corresponding elemental image behind the microlens stores the angular information from its perspective. From other point of view, considering the LF on the scene space, e.g. on the object plane of microlenses at d_c as shown

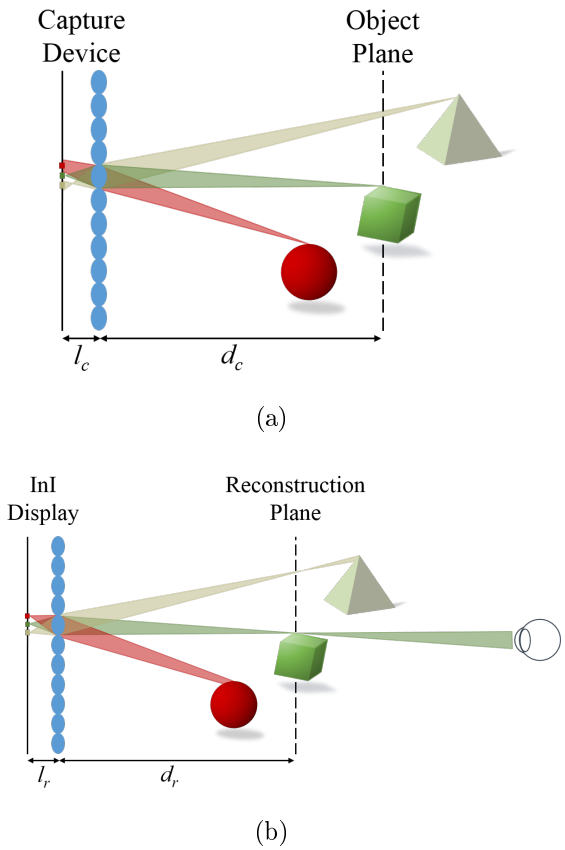


Figure 4.1 *LF (a) capture and (b) reconstruction of the given scene using array of cameras and InI display, respectively.*

in Fig. 4.1(a), each microlens samples angular information of the scene and the elemental images stores the spatial information.

In the display stage, the captured elemental images can be directly used to reconstruct the scene with the help of the microlens array in front of a 2D display as seen in Fig. 4.1(b). If the capture and the reconstruction parameters (i.e. pixel size, lens pitch, the gap between the lens and display planes) are not the same, the scaling should be handled correspondingly. To eliminate the cross-talk between the elemental images in both capture and reconstruction, physical barriers can be used in between each elemental image.

One of the main advantages of InI is that it is convenient to design a real-time capture-display system. However, for a real image projection, where the scene is formed in front of the lens array, the scene is reconstructed with reversed depth (pseudoscopic) as seen in Fig. 4.1(b). Various studies have addressed this problem. Okano et al. proposed a solution for a direct pickup system in [52], in which the display uses the virtual reconstruction method, as illustrated in Fig. 4.2. In this

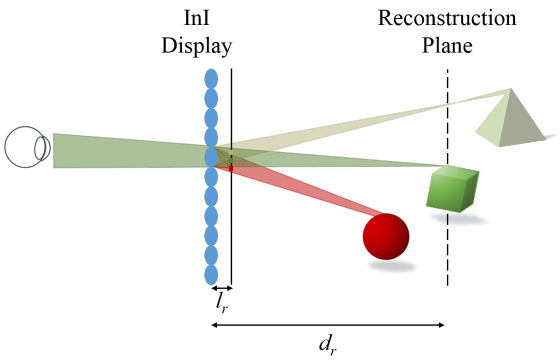


Figure 4.2 *Virtual scene reconstruction in InI display. The reconstructed scene is orthoscopic,*

technique, the elemental images are obtained by rotating the captured elemental images around their centers by 180°. Using the notations in Fig. 4.1(a), the distance between the microlens array and the display plane is set as $l_r = l_c - 2f^2(d_c - f)$, where f is the focal length of the microlenses. The virtual image is then obtained at $d_r = d_c - f$. A more generic algorithm is also presented in [49] to relate the pickup parameters to real or virtual display parameters. The technique proposed in [52] is, in fact, a special case of the algorithm formulated in [49].

InI reconstructs the LF of the scene by integrating the elemental images on the object position. As illustrated in Fig. 4.1(b), it actually integrates several beams focused by different microlenses and reconstructs focused points in space. From such points the light distributed continuously in different angles. By this way, the viewer is able to experience smooth motion parallax, when the head is moved. Moreover, due to providing such integrated set of narrow beams, accommodation cue can be also delivered. Thus, the viewer is able to focus at the scene points separated from the display surface, at which the eyes converge. Creation of correct accommodation cue is critical in the context of 3D displays. Because, otherwise, the viewer may experience visual fatigue due to mismatch between incorrect accommodation and vergence cues. This problem is known as accommodation-vergence conflict [24]. Various studies have demonstrated that InI is able deliver correct accommodation cue and thus avoid/reduce the accommodation-vergence conflict [23, 36, 32, 15].

The microlenses create blurred images at (out of focus) depths away from the reconstruction plane of the display. One way to eliminate the difference between the reconstruction quality, i.e. resolution, at the reconstruction plane and out of focus regions is to place the 2D display at the focal length of the microlenses, i.e. $l_r = f$ in Fig. 4.1(b) and Fig. 4.2, so as to obtain beams that propagate with (almost) same spatial extend (i.e. like in a cylindrical tube). In this way, the same resolu-

tion is achieved in a much larger depth range. However, as a trade-off, the overall spatial resolution is sacrificed [29, 45]. The detailed analysis of the viewing characteristics (spatial resolution, DoF, viewing angle etc.) of InI displays are presented in [11, 53, 46], and analyzed for both virtual and real projection type InI in [56]. Besides, in [27] the perceived resolution depending on the viewing plane is discussed and optimum display parameters are derived.

As discussed in Section 3.1.2, in the LF acquisition process of the focused plenoptic cameras the main lens maps the scene to be captured by the microlenses as perspective images. In that sense, the focused plenoptic camera can be considered as an InI capture setup with a main lens in front of the microlens array. As we will discuss later, similar correspondence exists between the defocused plenoptic camera and the SMV display. Such links between the capture and display techniques analyzed in this thesis are of critical importance in consideration of end-to-end 3D imaging/display systems.

More detailed information on InI, e.g. depth perception and resolution analysis, can be found in [69, 55].

4.2 Super Multiview Display

Multiview displays [59, 16, 25] are usually constructed by employing a lenticular array (cylindirical lens array) or a parallax barrier in front of a 2D conventional display e.g. LCD. Thus, unlike InI where 2D microlens array is used and full parallax imaging/display is achieved, MV displays usually provide horizontal parallax only (HPO) images. Please note, however, that generalization to full parallax case is possible, although not desired from practical considerations. Furthermore, instead of actually reconstructing the scene by focusing beams at the intended scene point as in InI, the MV displays direct slightly shifted (viewpoint) images of the scene to the viewer, who is assumed to be at the intended viewing plane. In the simplest case, two views can be generated for each eye of the viewer. However, usually the number views are higher e.g. 9. When the eyes are correctly positioned at the indented view location, they perceive slight shifted perspective images, where the shift amount (disparity) depends on the depth of the scene. By this way, binocular and vergence cues are delivered to the viewer.

One main drawback of the MV display is that since the views are constructed at quantized viewpoints, the viewer can observe jumps between adjacent views, which makes the parallax discontinuous, as opposed to the continuous parallax in InI. Besides, MV displays can not deliver accommodation cue. Therefore, the eyes of the

viewer focus on the display surface regardless of the position of the scene with respect to display, which results in suffering from accommodation-vergence conflict. SMV displays address both these issues by providing very dense set of views. In particular, the distance between adjacent viewpoints is kept smaller than the (average) eye pupil size so that each eye gets images from at least two adjacent views. This condition is called as the SMV condition [33]. By satisfying this condition the SMV displays are able to provide continuous (smooth) motion parallax as well as the accommodation-vergence conflict is avoided/reduced [61, 47]. Below, we analyze of the SMV display in terms of depth perception and resolution limitations.

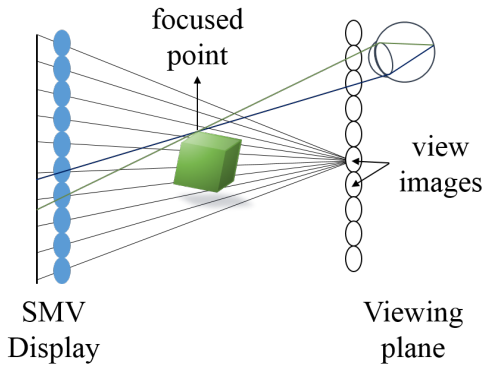


Figure 4.3 *Overall representation of an SMV display. At least two views are aimed within the eye pupil.*

The conceptual design of SMV display is given in Fig. 4.3. The generation of parallax images is such that the pitch between the neighborhood images, view pitch, is set to be smaller than the average eye pupil size ($\sim 5mm$). As illustrated in the figure, the eye receives at least two rays from a given scene point, which is necessary to create the correct accommodation cue. The eyes then focus on the scene point.

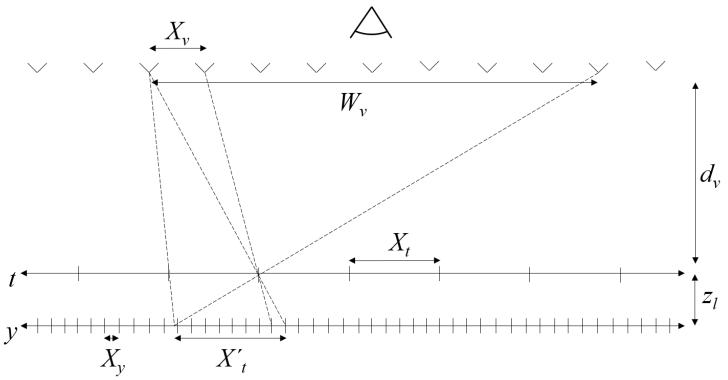


Figure 4.4 *SMV display design specification.*

Let us analyze the performance metrics of a SMV display for given display specifications. Device parametrization is illustrated in Fig. 4.4. The LCD and the lenticular planes are represented by y and t axes, respectively. The system is parametrized by the pixel pitch X_y of the LCD, lens pitch X_t and lens thickness z_l, all of which play roles in deriving the perceived resolution; the display DoF, the eyebox W_v (where the head can be moved freely), and the view pitch X_v that affects the depth perception. In the following, the related quality aspects of SMV display are discussed in details.

The SMV displays are periodic due to the periodicity in the lenticular lens, i.e. after the last view is seen, the viewer sees the first view again. This periodicity results in jumps on the border of each period. The eyebox of the display, W_v, is defined as one period of views. In this region the head can be moved freely. From the similar triangles the eyebox can be found as

$$W_v = \frac{X_t' d_v}{z_l} = X_t \frac{(d_v + z_l)}{z_l}, \tag{4.1}$$

where X_t', the multiplexing period, is slightly larger than X_t, i.e. $X_t' = X_t d_v / z_l$. It can then be concluded that for a fixed viewing distance d_v, the eyebox hinges upon the lens pitch. Within the eyeboxed derived, the number of views required for correct depth delivery and resulting perceived resolution is to be discussed in the following section.

4.2.1 Depth Perception in SMV Displays

Although they stimulate the vergence and binocular cues correctly, the MV displays suffer from the accommodation-vergence conflict and motion parallax issues, as described in the beginning of the section. In order to provide better visual experience, these issues should be properly addressed. For a SMV display, the view pitch X_v plays the key role in correct depth perception. Using the illustrations in Fig. 4.3 and Fig. 4.4, it can be concluded that the condition

$$X_v = \frac{X_y d_v}{z_l} \leq W_e, \tag{4.2}$$

has to be satisfied in order to utilize all depth cues, where W_e is the eye pupil size [61, 33].

Although X_v is the crucial parameter to evaluate the SMV display, in practice it is typical to express the device specification by the number of views within the eyebox.

By using Eq. 4.1 and Eq. 4.2, the number of views, N_v, can be derived as

$$N_v = \frac{W_v}{X_v} = \frac{X'_t}{X_y}. \tag{4.3}$$

Please note that Eq. 4.3 indicates the resemblance between the SMV displays and the defocused plenoptic cameras, in which the number of perspective images (i.e. number of views) is equal to the number of pixels inside an elemental image.

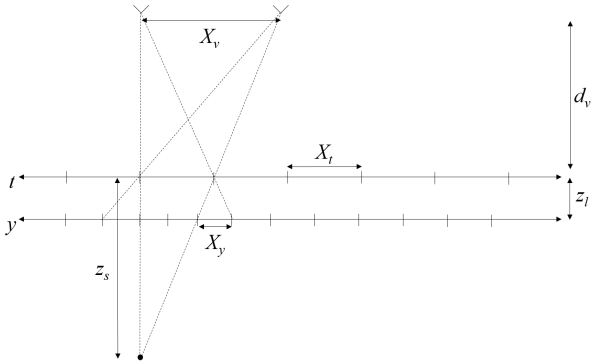


Figure 4.5 *The region where the smooth motion parallax is preserved when the full perceived resolution is utilized.*

The condition given by Eq. 4.2 ensures smoothness of the motion parallax for those objects that exhibit at most one pixel disparity between the adjacent view images [70]. Fig. 4.5 describes an object point seen from two adjacent views, of which the disparity is one perceived pixel. Beyond this object point the continuity of the motion parallax is not guaranteed. Indeed, it becomes dependent to the scene content. Therefore, the distance between the point and the display, z_s, can be described as the maximum distance that the smoothness of the parallax is ensured. Then, from similarity of triangles

$$z_s = -X_t \frac{d_v}{X_v - X_t}, \tag{4.4}$$

where minus sign indicates behind the display.

Although theoretical limit for smooth motion parallax can be derived by geometric analysis as given by Eq. 4.4, there are other factors in practice that also need to be taken into account. For instance, cross-talk between the view images makes the transition between the view images more smooth, thus works in the favor of smooth motion parallax. In fact, due to cross-talk the smooth motion parallax can be experienced even when the view pitch is larger than the eye pupil. This issue will be further investigated in Section 4.3 on a prototype SMV display.

4.2.2 Resolution Constraint

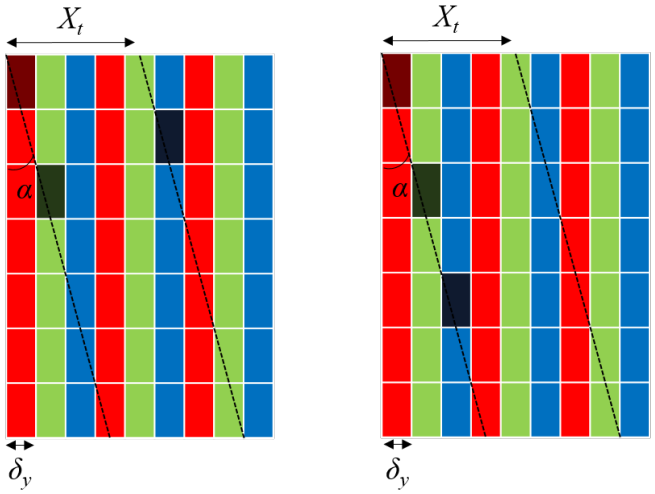
The above-mentioned advantages of the SMV displays require sacrifice from the perceived resolution, due to inherent spatio-angular trade-off. In particular, the increase in the spatial (texture) resolution results in the decrease of the angular resolution (number of views), and vice versa. The SMV displays achieve an accurate depth perception compared to MV by decreasing the view pitch (i.e. increasing the number of views in a fixed eyebox), but as a trade-off the perceived image resolution decreases. Therefore, the resolution limitation of SMV displays should be analyzed.

In the conventional SMV displays, where the center of projections of lenticular lenses are placed vertically, the perceived vertical resolution, N_{px}, is as much as the vertical resolution of the LCD, N_x. Perceived single view resolution in the horizontal direction, on the other hand, is equal to the LCD horizontal resolution, N_y, divided by the number of views N_v, $N_{py} = N_y/N_v$. Using Eq. 4.3, it can be concluded that

$$N_{py} = \frac{N_y X_y}{X'_t}. \tag{4.5}$$

For the SMV display where N_v is large, this drastic decrease in the horizontal resolution creates significantly uneven resolutions in horizontal and vertical directions. To overcome this problem, an elegant solution was proposed almost two decades ago in [63], by utilizing a subpixel level imaging technique. It suggests that instead of vertical placement, if the lenticular is placed with a specified slant angle, the subpixels belonging to different pixels merges at the viewing plane to construct one perceived (color) pixel of the same view. By this way the horizontal parallax image resolution is then increased at the cost of decrease in the vertical resolution. To illustrate the idea for better understanding, Fig. 4.6 is given, where two different subpixel combinations are drawn with the slant angle of the lenticular lens of $\alpha = \arctan(1/6)$. Please note that with the slanted array placement, X_t is not simply equal to the lens pitch T_l, instead $X_t = T_l/\cos(\alpha)$.

The subpixel arrangement in Fig. 4.6(a) illustrates that the view pixels are constructed by combining the subpixels from neighbor lenses and therefore the adjacent perceived pixels are located at every second lens. It can be concluded that the number of views are doubled in this configuration. Therefore, a non-slanted equivalent of the system can be constructed by defining $X_y = \delta_y/2$, i.e. the effective pixel pitch X_y is half of the subpixel pitch of the LCD δ_y. Doubling the number of views by decreasing the pixel pitch has a trad-off of decreasing the horizontal resolution to its half, according to Eq. 4.5. The perceived resolution in horizontal and vertical



(a) (b)

Figure 4.6 *Slanted lens arrangement in HPO SMV displays. By combining subpixels in different ways, the resolution loss in vertical direction can be by factor of (a) 3 and (b) 6.*

directions with the subpixel arrangement in Fig. 4.6(a) can then be calculated as

$$N'_{py} = 3\frac{N_y \delta_y}{2X'_t}, \tag{4.6}$$

$$N'_{px} = \frac{N_x}{3}, \tag{4.7}$$

and using Fig. 4.6(b)

$$N'_{py} = 6\frac{N_y \delta_y}{2X'_t}, \tag{4.8}$$

$$N'_{px} = \frac{N_x}{6}, \tag{4.9}$$

respectively. Fig. 4.6(b) is preferable when horizontal resolution drops drastically due to large number of views.

4.2.3 Display Depth of Field

The theoretical analysis of the display resolution is consistent as long as the maximum extent of the scene is within a certain limit, which is defined as the display DoF. Behind the DoF, the resolution of the content should be limited accordingly, in order to prevent aliasing [73]. Below, the spatio-angular sampling in SMV display, display bandwidth and DoF are analyzed in order to better understand the resolution constraints for a given scene.

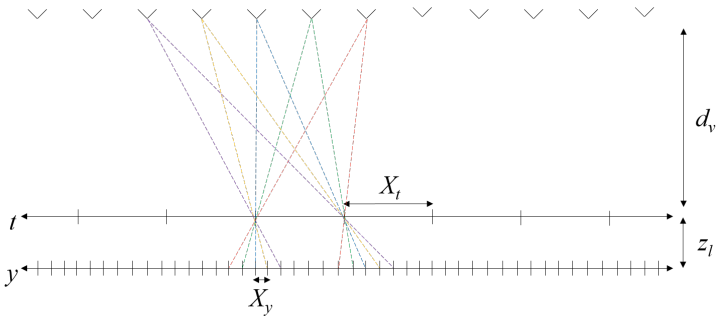


Figure 4.7 *Sampling analysis of a SMV display.*

The propagation of rays from different views are given with the display parameters in Fig. 4.7. The sampling grid of the system can then be represented as in Fig. 4.8(a). Please note that the coordinates of the rays in $y-axis$ is relative to the $t-coordinates$ of the lenses, i.e. each view coordinate in $y-axis$ is defined around its corresponding lens coordinate in t. Due to the slight difference between X_t' and X_t, the shear effect occurs in the sampling grid, as seen in Fig. 4.8(a). The slope of shear is $s = z_l/d_v$.

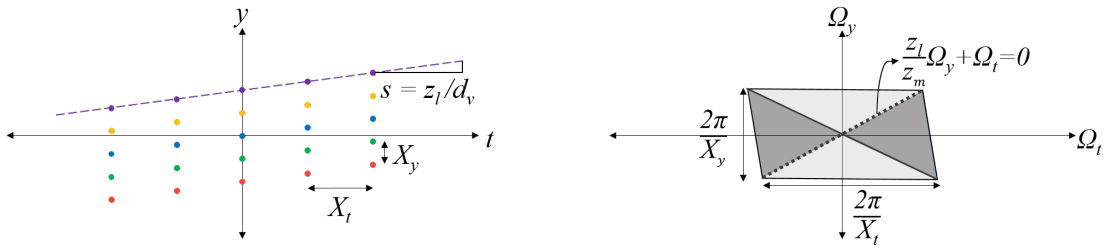


Figure 4.8 *Sampling analysis of SMV display, (a) ray-space sampling grid, (b) display bandwidth in Fourier domain.*

Defining the Fourier transform of $f(t, y)$ as $F(\Omega_t, \Omega_y)$, the frequency analysis of the vertically sheared function $f(t, y + st)$ gives

$$F\{f(t, y + st)\} = F(\Omega_t - s\Omega_y, \Omega_y), \tag{4.10}$$

meaning that the vertical shear in spatial domain introduces horizontal shear in the frequency domain [8]. Therefore, the frequency response of the display with given sampling rate results in the parallelogram bandwidth as in Fig 4.8(b).

The continuous light field of a scene with constant depth z_0 is represented by the line $\Omega_y z_l/z_0 + \Omega_t = 0$; furthermore, the bandwidth of the continuous light field of a

scene in depth range $[z_1, z_2]$ is described to be in between the lines $\Omega_y z_l/z_1 + \Omega_t = 0$ and $\Omega_y z_l/z_2 + \Omega_t = 0$ [12]. The reconstruction at full spatial resolution without aliasing is then possible, if the boundaries of the scene lies within the dark gray region in Fig. 4.8(b). The diagonal lines bounding this region correspond the DoF boundaries of the display, where the solid line represents in front of and the dashed one represents behind the display. The boundary behind the display can be found as

$$z_m = -z_l \frac{X_t}{X_y - sX_t}. \tag{4.11}$$

Replacing s with z_l/d_v, the equation can be rearranged as

$$z_m = -\frac{d_v X_t}{X_y \frac{d_v}{z_l} - X_t}. \tag{4.12}$$

Eq. 4.12 and Eq. 4.4 reveal that the DoF limit of the display is equal to the maximum distance that guarantee the smooth motion parallax, i.e. $z_s = z_m$. As discussed in [73] and also seen in Fig. 4.8(b), for a scene at $|z| > |z_m|$ behind the display, the content resolution should be decreased by the factor of z_m/z with respect to the maximum available (perceived) resolution in order to prevent aliasing in reconstruction.

Similar to the relation between the focused plenoptic camera and the InI device explained in Section 4.1, a correspondence between the defocused plenoptic camera and the SMV display can be constructed as well, assuming the full parallax extension of the (HPO) design. The capture setup of the defocused plenoptic camera is designed to obtain the view images at the main lens plane, while in the SMV display these view images are constructed at the viewing plane, as discussed in details.

4.3 SMV-HUD Prototype System

In this section we present a head up display (HUD) system prototype that we develop based on the SMV display technique. HUD can be defined as a semi-transparent display that enables driver viewing the necessary information via its simulated image superposed on top of the real world scene. By this way, unlike in the case of conventional instrument clusters, the driver is always able to keep his/her eyes on the road and thus a safer driving condition is satisfied [35, 34]. First HUD designs were mainly designed to be used in military aircrafts [30]. However, for more than three decades, the commercial applications of HUD have been available as well. The General Motors developed the first HUD for automotive industry in 1988, which consists of the vacuum fluorescent tube, reflective optics and the windshield [65].

The automotive HUD applications can be categorized into two as the direct and virtual projection HUDs. In a conventional direct projection HUD [22], the information is presented on top of the semi-transparent reflector, such as windshield. Although this technique provides a feasible HUD solution, it does not properly address the reaccommodation problem, i.e. the driver needs to continuously refocus between the real world and the displayed image. Virtual projection HUDs solve this problem by projecting the images behind the windshield within a reasonable accommodation distance ($2-3m$), utilizing several display techniques [4, 48, 10, 13]. Most of the existing virtual-HUDs, however, require extra installation space to achieve the desired virtual depth ranges, or able to provide limited size images [4, 10, 13]. The SMV display technique provides a compact, scalable and cheap alternative solution for virtual projection HUD systems, since it basically consist of lens array and a LCD, which together at most few centimeters thick and the scale of the presented image is simply dependent on the LCD (and the lens array) size. Please note that a SMV based windshield display has been previously introduced in [60]. In their implementation, however, relatively large distances has been targeted and thus a Fresnel lens has to be employed over the display, which results in at least few tens of centimeters thick form-factor. Below, the detailed theoretical analysis of the proposed solution is presented together with corresponding experimental verification.

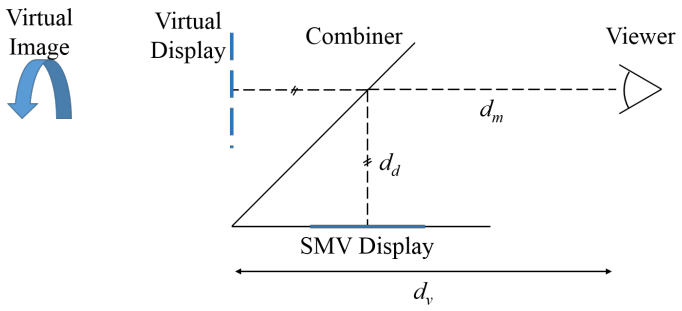


Figure 4.9 *The overall representation of the SMV-HUD system.*

The overall representation of the proposed SMV-HUD system is illustrated in Fig. 4.9. A typical HUD system has an image preparation unit to generate the information for the driver, and a combiner (e.g. windshield) to augment the generated image to the real world outside. For the proposed SMV-HUD prototype, the virtual display plane, i.e. the plane at which the display is reflected by the combiner, and the viewing plane constitute the SMV display parametrization discussed in Fig. 4.4. The viewing distance d_v in the simplified system is equal to the distance between the viewer and the combiner, d_m, plus the distance between the combiner and the display, d_d.

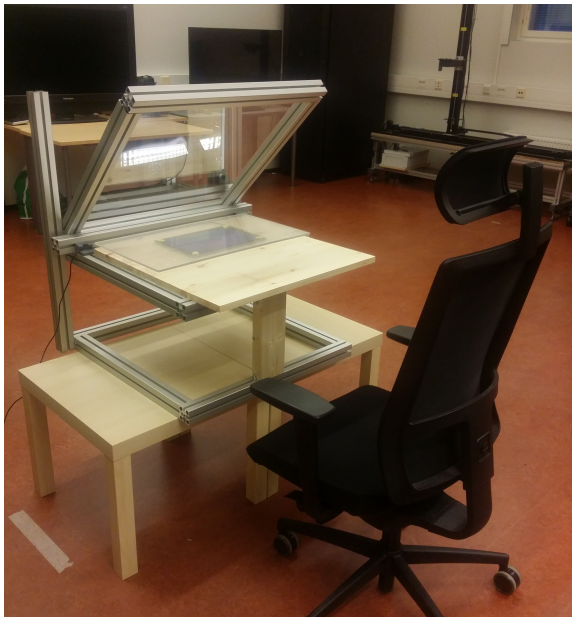


Figure 4.10 *The prototype of SMV-HUD.*

The developed SMV-HUD prototype is given in Fig. 4.10. The device specification of the SMV display are the following: the lens pitch is $T_l = 1.57mm$, the slant angle is $\alpha = \arctan(1/6)$, which makes the equivalent system lens pitch $X_t = 1.59mm$, the LCD size is $16.9cm \times 25.4cm$ with subpixel pitch $\delta_y = 39\mu m$ in the horizontal and $\delta_x = 117\mu m$ in the vertical directions. The equivalent (effective) horizontal pixel pitch X_y is the half of δ_y, therefore $X_y = 19.5\mu m$. The vertical equivalent pixel pitch X_x is the same as δ_x, i.e. $X_x = 117\mu m$. The lens thickness is $z_l = 4.25mm$ and the viewing distance is $d_v = 90cm$. With the proposed system parameters, resulting eyebox is $W_v = 33.80cm$, and there are 82 views within the eyebox in which the interview distance is $X_v = 4.14mm$.

The derived system capabilities are typical for the SMV display, regardless of the application. For HUD system qualification, field of view (FOV) for a fixed d_v is also an effective parameter to evaluate. The FOV of the SMV-HUD can be calculated using the display width and height. In particular, the FOV is

$$FOV_y = 2\arctan\Big(\frac{W_y}{2d_v}\Big), FOV_x = 2\arctan\Big(\frac{W_x}{2d_v}\Big), \tag{4.13}$$

in the horizontal and vertical directions, respectively. W_x is the height ($16.9cm$) and W_y is the width ($25.4cm$) of the display. Therefore, $FOV_y = 16.04°$ and $FOV_x = 10.73°$.

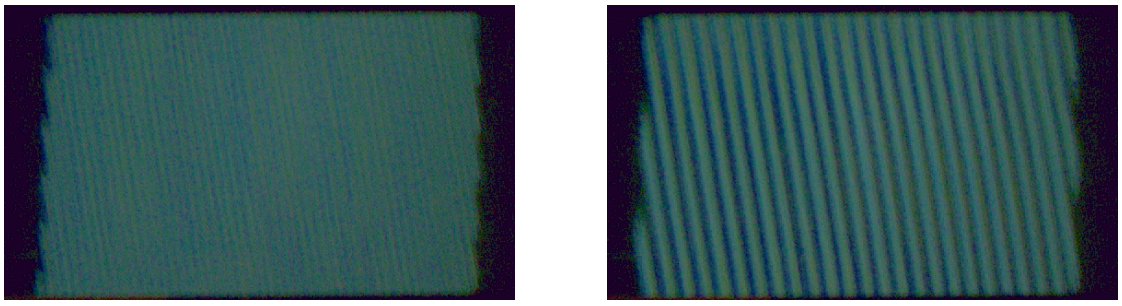
With the subpixel mapping in Fig. 4.6(b), the full (maximum available) perceived

resolution is 240×160 pixels. As discussed in Section 4.2.3, this resolution is available for the scenes inside the display DoF and for those beyond the DoF, the resolution should be reduced accordingly. Using Eq. 4.12 with given parameters, the maximum extent of the DoF is calculated as $z_m = -55.95cm$. If the virtual image is aimed e.g. at $2.5m$ from the driver, which is typical for existing virtual-HUDs, the scene should be reconstructed at $1.6m$ behind the display ($z = -1.6m$), for $d_v = 90cm$. Therefore, the horizontal perceived resolution decreases by factor of $z_m/z \approx 1/3$ with respect to maximum available resolution.

Below, the quality factors of the implemented prototype are analyzed and the system capacity is examined. The depth and resolution perception are experimented with the help of a precise rig system on which 1920 × 1200 pixels resolution camera is mounted. The distance between the camera and the virtual display plane is assured to be $d_v = 90cm$.

4.3.1 Perceived Resolution

The theoretical resolution constraints are tested at different depth values utilizing sinusoidal signals with different frequencies. The boundary of the display DoF, $1.45m$ from the viewer, and beyond that value, $3m$, are chosen as the test depths.

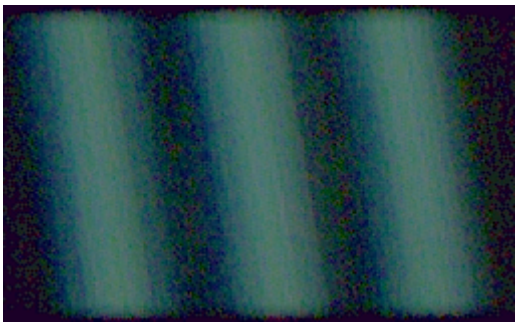


(a) (b)

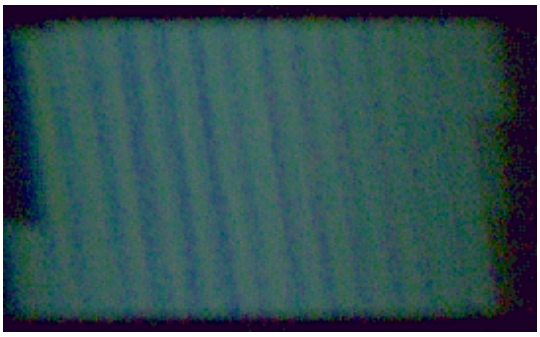
Figure 4.11 *Resolution experiment for the image distance of* $1.45m$ *away from the viewer. (a) Maximum resolution, (b) half of maximum resolution.*

In the first phase of the resolution experiments, two different sinusoidal images (rotated by the slant angle) are to be reconstructed by the display. Since the images are placed at $1.45m$ from the camera, they are $55cm$ behind the display, which is within the display bandwidth. Therefore, the display is expected to reconstruct the pattern with the full perceived resolution. Fig. 4.11(a) illustrates the sinusoidal pattern with the maximum frequency, where the period is two perceived pixel, so that each line corresponds to one lens. In Fig. 4.11(b), the reconstructed pattern

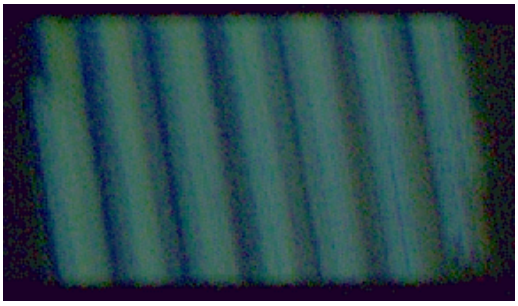
has half of the maximum frequency. It can be observed that the display reconstructs both patterns correctly.



(a)



(b)



(c)

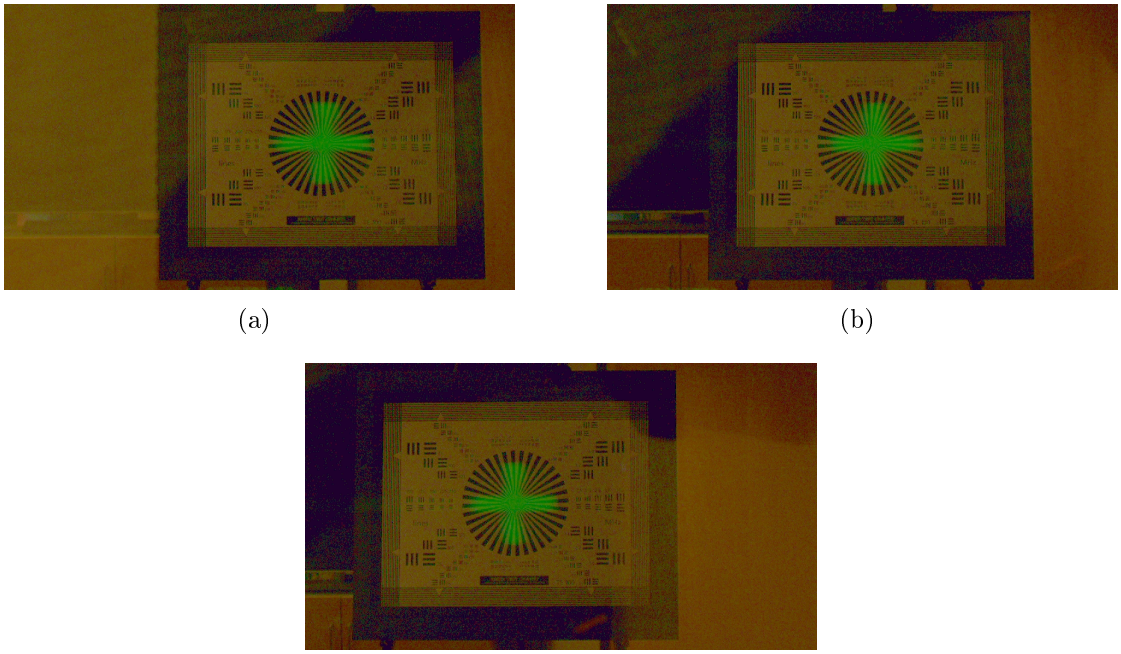
Figure 4.12 *Resolution experiment for the image distance of* $3m$ *away from the viewer. (a) Maximum perceived resolution within DoF, (b) reduced resolution for* $3m$*, (c) half of the reduced resolution.*

The second examination of the resolution limits are done outside the display DoF. The reconstruction of three sinusoidal contents at $3m$ are observed, as illustrated in Fig. 4.12. The pattern with the frequency corresponding to the maximum resolution, shown in Fig. 4.12(a), is aliased, while the signal carrying the frequency decreased by the factor of z/z_m ($2.1m/0.56m$) in Fig. 4.12(b) and the one with half of this reduced resolution in Fig. 4.12(c) are reconstructed as expected. Therefore, the theoretical analysis of the resolution presented in Section 4.2.2 is verified in

practice.

4.3.2 Perceived Depth

The importance of delivering correct depth cues have been discussed in Section 4.2.1. The depth perception evaluation constitutes the second phase of the experiments with the SMV-HUD prototype. Mainly the binocular cues and the smoothness of the motion parallax are tested. The cross-talk between the views and its effect to motion parallax are discussed.



(a) (b)

(c)

Figure 4.13 *Binocular disparity experiment utilizing the views at (a) 10cm left, (b) 0, and (c) 10cm right with respect to the middle view. The test chart is placed at* $3m$ *from the camera.*

The correctness of the binocular disparity is evaluated by utilizing a test chart placed at the intended virtual image depth, $3m$ away from the camera. The motion of the virtual image is then observed with respect to the test chart. The results are presented in Fig. 4.13, where three different views are captured from middle view and from views $10cm$ left and right to the middle view. The images shows that the position of the virtual image with respect to the test chart does not change, i.e. they deliver the same parallax with changing horizontal position of the camera. It can then be concluded that the virtual image exhibit the same disparity value with the test chart and the view images satisfy the binocular cue at the target image depth.

Please note that because of various manufacturing issues, e.g. the lenticular is not exactly flat, or the lens pitch is not exactly same for each lens, the virtual image size and position is distorted in negligible amount.

Motion parallax is a critical cue for depth perception for the intended depth range. In order to promote the advantages of HUD, e.g. more comfortable and safer driving, the smoothness of the motion parallax should be provided. The implemented prototype has the view pitch X_v smaller than the eye pupil size, in which case it is expected that the smooth motion parallax is ensured inside the DoF of the display. On the other hand, for the virtual images beyond the DoF, in theory the image jumps are possible during view transition depending on the image content. As discussed in Section 4.2.1, however, cross-talk may extend the theoretical limit by enlarging the visible area of each view. Thus, before the motion parallax experiments, it is beneficial to evaluate the cross-talk.

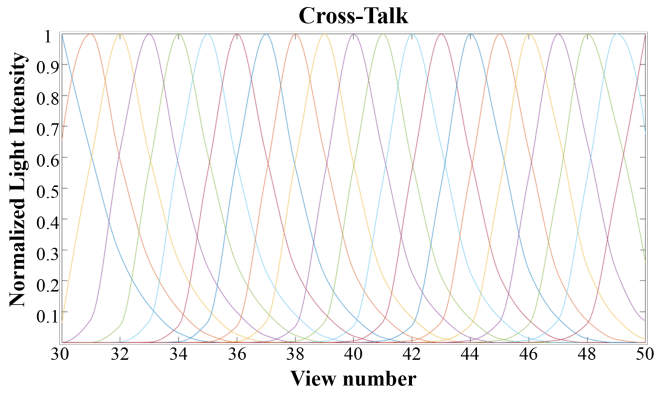
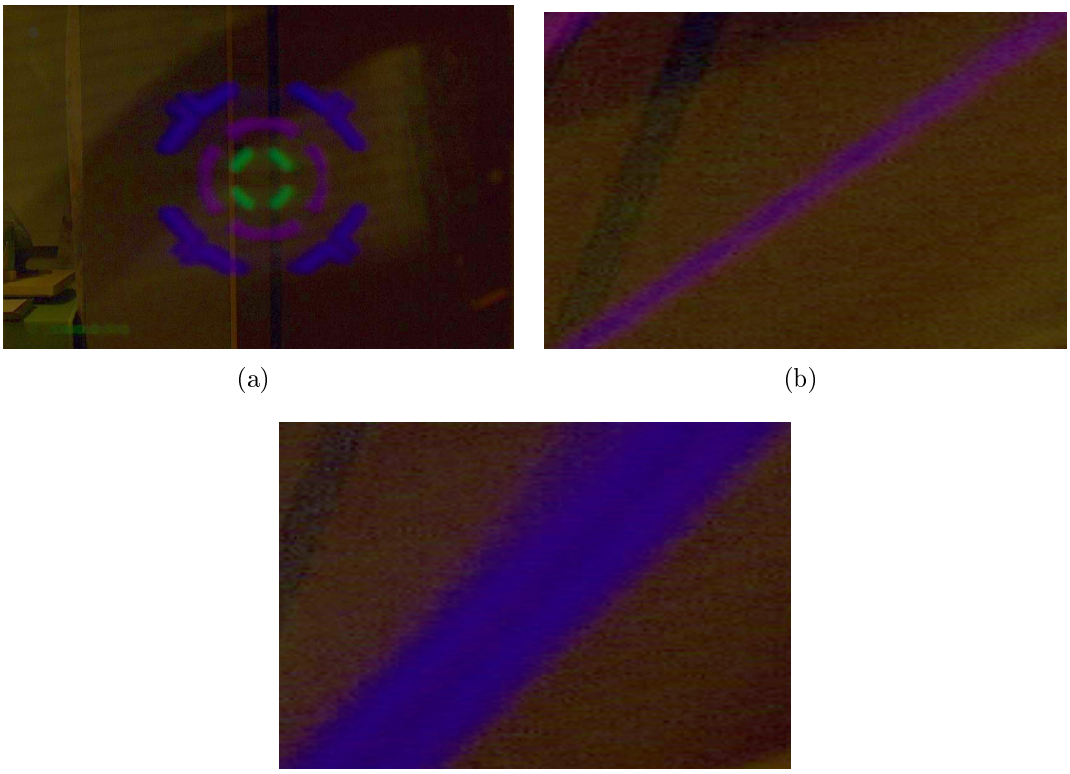


Figure 4.14 *Cross-talk, the interaction between the views.*

The cross-talk values are obtained by measuring the average intensity of all views at each view locations. For each measurement then, all the subpixels belonging to the corresponding view have the maximum intensity value (i.e. a white image is constructed) while all other views are black. The camera captures the images from the center of each view and the average intensities are stored. The interactions between the views $[30, 50]$ are illustrated in Fig. 4.14. The cross-talk can be defined as the proportion of the leakage luminance to the intended luminance at the corresponding viewpoint, $c = leakage/signal * 100\%$ [68]. Using this definition, the average cross-talk is calculated to be 207.7%. Thus, there is a significant amount of cross-talk between the views that plays role in smoothing the view transitions.

In order to test the motion parallax, three circles at varying depth values are introduced to the display, the innermost circle is at $1.5m$, the middle one is at $2m$, and the outermost circle is placed at $3m$. For visualization of the view transition, the



(a) (b)

(c)

Figure 4.15 *Smooth motion parallax test for a 3D scene consisting of different depth planes at* $1.5m$*,* $2m$*, and* $3m$*. (a) Captured image from the middle view, (b)-(c) two example epipolar plane images corresponding to depth planes at* $2m$ *and* $3m$*, respectively.*

camera is moved horizontally in the eyebox to capture the view images, from which the same vertical positions (i.e. rows of the images) are extracted and the epipolar plane images (EPI) are constructed by stacking these rows in the vertical direction. The EPIs belong to the middle circle ($2m$) in Fig. 4.15(b) and the furthest circle ($3m$) in Fig. 4.15(c). From the continuity of the epipolar lines of the circles, it can be concluded that the view transition is perceived as continuous. The significant amount of cross-talk calculated above has also a role in the continuity of the motion parallax.

Although the experimental results are coherent with the theoretical analysis and give better understanding of the subject, please note that the presented implementation does not utilize the best parameters possible. Instead, by utilizing the commercially available products, a proof of the concept is aimed. Thus, the display can be developed further with the existing technology, e.g. by using QLED with denser subpixel resolution as the 2D display, or by using a custom design lenticular lens with the parameters optimized for the HUD application.

5. CONCLUSIONS

Lens array based 3D capture and display devices construct the conceptual framework of this thesis. Defocused and focused plenoptic cameras are investigated in the scope of the capture systems, while the InI and SMV displays are studied as lens array based 3D display techniques. Besides the theoretical analysis, the defocused plenoptic camera is simulated utilizing a computer graphics rendering algorithm. Such a tool enables, for instance, to obtain ground-truth plenoptic data for an animated scene. Moreover, a HUD prototype has been presented as an attractive application of the SMV display technique.

As addressed in the theoretical analysis, there is actually a link between the defocused plenoptic camera and the SMV display (in its full parallax extension), and between the focused plenoptic camera and the InI setup. Such conceptual similarity can be useful, for instance, in a real-time broadcasting system where the 3D content can be captured e.g. by the defocused plenoptic camera and delivered to the viewer by an SMV display.

The spatio-angular trade-off inherent to plenoptic cameras indicate that the spatial (texture) resolution can be enhanced at the expense of decrease in the angular resolution. In the theoretical analysis of the plenoptic cameras in Chapter 3, we have demonstrated that the focused plenoptic camera favors spatial information, whereas the angular resolution is desired to be kept sufficiently high in the defocused plenoptic camera. Therefore, both cameras can be superior to each other depending on the post-processing application to be implemented. For the applications where denser depth planes increase the performance, e.g. refocusing, angular information is of more importance, therefore the defocused plenoptic camera can be utilized. Similarly, for the applications that requires higher spatial resolution, e.g. depth estimation, focused plenoptic camera can provide better results.

The simulation algorithm implemented for the defocused plenoptic camera utilizes the densely sampled light field framework, which enables defining the pixel intensities in terms of explicit accurate integrals rather than approximations based on ray oversampling. Having the ground-truth plenoptic image, one can reliably test the

performance of a desired post-processing algorithm, e.g. refocusing. In the current implementation the microlenses are assumed to be of square shape. As a future work, different microlens arrangements such as circular or hexagonal are also considered to be implemented.

The discussion of InI and the SMV displays have also revealed the spatio-angular trade-offs in such displays. The angular resolution in the InI can be enhanced by increasing the number of microlenses as a trade-off of decrease in the spatial resolution, as in the focused plenoptic camera. On the other hand, the increase in the number of lenses in the SMV display results in enhancement of the spatial resolution as a trade-off of decrease in the angular resolution (i.e. number of views), as in the defocused plenoptic camera.

Regarding the practical implementation of lens array based 3D displays, the SMV display technique has been utilized to develop a SMV-HUD prototype. The prototype has demonstrated an attractive solution for the virtual image presentation HUD application. In particular, the image can be presented at the intended virtual depths $2-3m$ with correct depth cues. In doing this, unlike most of the existing virtual-HUDs, the SMV-HUD does not require an extra installation space. Furthermore it is almost arbitrarily scalable depending on the size of the display.

As also addressed in the theoretical analysis, the development process of the prototype SMV display presented in Section 4.2 has demonstrated that the parameter selection is an optimization problem and the device specifications should be decided in accordance with the requirements of the application. In the HUD prototype where the virtual image distance is aimed as much as $2-3m$, for instance, the display DoF is kept to be rather large, which then results in sacrifice from the resolution. For another application where the scene is around the display plane, the DoF can be kept narrow, and resolution can be increased. Similarly, for applications requiring smaller eyebox, the resolution can be enhanced more. Such trade-offs should be carefully considered in order to provide better user experience. Please also note that the SMV-HUD system has been developed as a proof of concept. Thus, moderate resolution LCD is used with a lenticular sheet available in the market. Thus, the presented system can be significantly enhanced by utilizing custom high-quality lens arrays and higher resolution displays.

BIBLIOGRAPHY

[1] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," 1991.

[2] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, Feb. 1992.

[3] U. Akpinar, E. Sahin, and A. Gotchev, "Simulation of microlens array based plenoptic capture utilizing densely sampled light field," in *Proc. 3DTV Conference 2017 (3DTV-CON 2017)*, Copenhagen, Denmark, June 2017.

[4] J. R. Banbury, "Head-up display systems," *Science Progress (1933-)*, pp. 497–517, 1992.

[5] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 972–986, 2012.

[6] B. Blundell, *Enhanced Visualization: Making Space for 3-D Images*. Wiley, 2007.

[7] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.

[8] R. Bracewell, *The Two-Dimensional Fourier Transform*. Boston, MA: Springer US, 2003, pp. 157–159. [Online]. Available: https://doi.org/10.1007/978-1-4419-8963-5_4

[9] D. Brewster, *The Stereoscope; its History, Theory, and Construction, with its Application to the fine and useful Arts and to Education: With fifty wood Engravings*. John Murray, 1856.

[10] E. Buckley and D. Stindt, "Full colour holographic laser projector hud," in *Proceedings: SID Annu. Symp. Veh. Displays*, vol. 15, 2008, pp. 131–135.

[11] C. B. Burckhardt, "Optimum parameters and resolution limitation of integral photography," *J. Opt. Soc. Am.*, vol. 58, no. 1, pp. 71–76, Jan 1968. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=josa-58-1-71

[12] J. Chai, X. Tong, S. Chan, and H. Shum, "Plenoptic sampling." ACM, July 2000, pp. 307–318. [Online]. Available: https://www.microsoft.com/en-us/research/publication/plenoptic-sampling/

[13] J. Christmas, D. Masiyano, and N. Collings, "Holographic automotive head up displays," *Electronic Displays Conference*, Feb 2015.

[14] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1027–1034.

[15] H. Deng, Q.-H. Wang, C.-G. Luo, C.-L. Liu, and C. Li, "Accommodation and convergence in integral imaging 3D display," *Journal of the Society for Information Display*, vol. 22, no. 3, pp. 158–162, 2014.

[16] N. A. Dodgson, J. R. Moore, and S. R. Lang, "Multi-view autostereoscopic 3D display."

[17] J. Geng, "Three-dimensional display technologies," *Advances in optics and photonics*, vol. 5, no. 4, pp. 456–535, 2013.

[18] T. Georgiev, G. Chunev, and A. Lumsdaine, "Superresolution with the focused plenoptic camera," in *Proc. SPIE*, vol. 7873, 2011, pp. 78 730X–78 730X–13. [Online]. Available: http://dx.doi.org/10.1117/12.872666

[19] T. Georgiev and A. Lumsdaine, "Superresolution with plenoptic camera 2.0."

[20] ——, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 021 106–021 106, 2010.

[21] A. Gershun, "The light field," *Studies in Applied Mathematics*, vol. 18, no. 1-4, pp. 51–151, 1939.

[22] M. K. Hedili, M. O. Freeman, and H. Urey, "Microlens array-based high-gain screen design for direct projection head-up displays," *Appl. Opt.*, vol. 52, no. 6, pp. 1351–1357, Feb 2013. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-52-6-1351

[23] H. Hiura, T. Mishina, J. Arai, and Y. Iwadate, "Accommodation response measurements for integral 3D image."

[24] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision*, vol. 8, no. 3, pp. 33–33, 2008.

[25] N. S. Holliman, N. A. Dodgson, G. E. Favalora, and L. Pockett, "Three-dimensional displays: a review and applications analysis," *IEEE transactions on Broadcasting*, vol. 57, no. 2, pp. 362–371, 2011.

[26] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, "Three-dimensional display technologies of recent interest: principles, status, and issues [invited]," *Appl. Opt*, vol. 50, no. 34, pp. H87–H115, 2011.

[27] H. Hoshino, F. Okano, H. Isono, and I. Yuyama, "Analysis of resolution limitation of integral photography," *J. Opt. Soc. Am. A*, vol. 15, no. 8, pp. 2059–2065, Aug 1998. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-15-8-2059

[28] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 297–306.

[29] J.-S. Jang, F. Jin, and B. Javidi, "Three-dimensional integral imaging with large depth of focus by use of real and virtual image fields," *Opt. Lett.*, vol. 28, no. 16, pp. 1421–1423, Aug 2003.

[30] D. N. Jarrett, *Cockpit Engineering*. Ashgate Pub., 2005. [Online]. Available: https://books.google.fi/books?id=WKhTAAAAMAAJ

[31] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß, *On the Calibration of Focused Plenoptic Cameras*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 302–317. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-44964-2_15

[32] J.-H. Jung, K. Hong, and B. Lee, "65.2: Effect of viewing region satisfying super multi-view condition in integral imaging," in *SID Symposium Digest of Technical Papers*, vol. 43, no. 1. Wiley Online Library, 2012, pp. 883–886.

[33] Y. Kajiki, H. Yoshikawa, and T. Honda, "Hologramlike video images by 45-view stereoscopic display," vol. 3012, 1997, pp. 154–166.

[34] R. Kiefer, "Quantifying head-up display (hud) pedestrian detection benefits for older drivers," in *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles*, 1998, pp. 428–437.

[35] R. Kiefer and A. Gellatly, "Quantifying the consequences of the 'eyes-on-road' benefit attributed to head-up displays," in *SAE Technical Paper*. SAE International, 02 1996. [Online]. Available: http://dx.doi.org/10.4271/960946

[36] Y. Kim, J. Kim, K. Hong, H. K. Yang, J. H. Jung, H. Choi, S. W. Min, J. M. Seo, J. M. Hwang, and B. Lee, "Accommodative response of integral imaging in near distance," *Journal of Display Technology*, vol. 8, no. 2, pp. 70–78, Feb 2012.

[37] R. A. Lessard and H. I. Bjelkhagen, "Practical holography XX: Materials and applications." SPIE, 2006.

[38] M. Levoy, "Light fields and computational imaging," *Computer*, vol. 39, no. 8, pp. 46–55, 2006.

[39] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.

[40] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 924–934, 2006.

[41] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision*, vol. 58, no. 2, pp. 121–138, 2004.

[42] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908.

[43] B. Liu, Y. Yuan, S. Li, Y. Shuai, and H.-P. Tan, "Simulation of light-field camera imaging based on ray splitting Monte Carlo method," *Optics Communications*, vol. 355, pp. 15 – 26, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0030401815005283

[44] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *2009 IEEE International Conference on Computational Photography (ICCP)*, April 2009, pp. 1–8.

[45] S.-W. Min, B. Javidi, and B. Lee, "Enhanced three-dimensional integral imaging system by use of double display devices," *Appl. Opt.*, vol. 42, no. 20, pp. 4186–4195, Jul 2003.

[46] S.-W. Min, J. Kim, and B. Lee, "New characteristic equation of three-dimensional integral imaging system and its applications," *Japanese journal of applied physics*, vol. 44, no. 1L, p. L71, 2004.

[47] H. Mizushina, J. Nakamura, Y. Takaki, and H. Ando, "Super multi-view 3D displays reduce conflict between accommodative and vergence responses," *Journal of the Society for Information Display*, vol. 24, no. 12, pp. 747–756, 2016, sID-08-16-0434.R1. [Online]. Available: http://dx.doi.org/10.1002/jsid.520

[48] K. Nakamura, J. Inada, M. Kakizaki, T. Fujikawa, S. Kasiwada, H. Ando, and N. Kawahara, "Windshield display for intelligent transport system," in *11th World Congress on Intelligent Transportation Systems*, vol. 85, 2004, pp. 3058–3065.

[49] H. Navarro, R. Martínez-Cuenca, G. Saavedra, M. Martínez-Corral, and B. Javidi, "3D integral imaging display by smart pseudoscopic-to-orthoscopic conversion (spoc)," *Opt. Express*, vol. 18, no. 25, pp. 25 573–25 583, Dec 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-25-25573

[50] R. Ng, *Digital light field photography*. Stanford University California.

[51] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.

[52] F. Okano, H. Hoshino, J. Arai, and I. Yuyama, "Real-time pickup method for a three-dimensional image based on integral photography," *Applied optics*, vol. 36, no. 7, pp. 1598–1603, 1997.

[53] T. Okoshi, "Optimum design and depth resolution of lens-sheet and projection-type three-dimensional displays," *Appl. Opt.*, vol. 10, no. 10, pp. 2284–2291, Oct 1971. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-10-10-2284

[54] ——, *Three-dimensional imaging techniques*. Elsevier, 2012.

[55] J.-H. Park, K. Hong, and B. Lee, "Recent progress in three-dimensional information processing based on integral imaging," *Appl. Opt.*, vol. 48, no. 34, pp. H77–H94, Dec 2009.

[56] J.-H. Park, S.-W. Min, S. Jung, and B. Lee, "Analysis of viewing parameters for two display methods based on integral photography," *Appl. Opt.*, vol. 40, no. 29, pp. 5217–5232, Oct 2001. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-40-29-5217

[57] A. Razavi, "Simulating the plenoptic camera." 2003.

[58] E. Sahin, V. Katkovnik, and A. Gotchev, "Super-resolution in a defocused plenoptic camera: a wave-optics-based approach," *Opt. Lett.*, vol. 41, no. 5, pp. 998–1001, Mar 2016.

[59] J.-Y. Son, B. Chae, W.-H. Son, J. Nam, and B.-R. Lee, "Comparisons of viewing zone characteristics of multiview and integral photography 3D imaging methods," *Journal of Display Technology*, vol. 8, no. 8, pp. 464–471, 2012.

[60] Y. Takaki, Y. Urano, S. Kashiwada, H. Ando, and K. Nakamura, "Super multi-view windshield display for long-distance image information presentation," *Opt. Express*, vol. 19, no. 2, pp. 704–716, Jan 2011. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-19-2-704

[61] Y. Takaki, "Development of super multi-view displays," *ITE Transactions on Media Technology and Applications*, vol. 2, no. 1, pp. 8–14, 2014.

[62] I. Tosic, S. A. Shroff, and K. Berkner, "Dictionary learning for incoherent sampling with application to plenoptic imaging," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 1821–1825.

[63] C. van Berkel, "Image preparation for 3D LCD," in *Proc. SPIE*, vol. 3639, 1999, pp. 84–91. [Online]. Available: http://dx.doi.org/10.1117/12.349368

[64] N. J. Wade, H. Ono, and L. Lillakas, "Leonardo da vinci's struggles with representations of reality," *Leonardo*, vol. 34, no. 3, pp. 231–235, 2001.

[65] M. Weihrauch, G. Meloeny, and T. Goesch, "The first head up display introduced by general motors," SAE Technical Paper, Tech. Rep., 1989.

[66] C. Wheatstone, "Contributions to the physiology of vision.–part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical transactions of the Royal Society of London*, pp. 371–394, 1838.

[67] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 765–776.

[68] A. J. Woods and C. R. Harris, "Comparing levels of crosstalk with red/cyan, blue/yellow, and green/magenta anaglyph 3D glasses," 2010.

[69] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: sensing, display, and applications," *Applied optics*, vol. 52, no. 4, pp. 546–560, 2013.

[70] M. Yamaguchi, "Light-field and holographic three-dimensional displays," *JOSA A*, vol. 33, no. 12, pp. 2348–2364, 2016.

[71] C. Zhang and T. Chen, "Spectral analysis for sampling image-based rendering data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1038–1050, 2003.

[72] R. Zhang, P. Liu, D. Liu, and G. Su, "Reconstruction of refocusing and all-in-focus images based on forward simulation model of plenoptic camera," *Optics Communications*, vol. 357, pp. 1–6, 2015.

[73] M. Zwicker, W. Matusik, F. Durand, and H. Pfister, "Antialiasing for automultiscopic 3D displays," in *ACM SIGGRAPH 2006 Sketches*. ACM, 2006, p. 107.