



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학박사 학위논문

**Image processing methods for light-field
based computational imaging**

라이트-필드 기반의 전산 이미징을
위한 연산 처리 방법

2019년 8월

서울대학교 대학원

전기컴퓨터공학부

정 영 모

Image processing methods for light-field based computational imaging

지도 교수 이 병 호

이 논문을 공학박사 학위논문으로 제출함
2019년 8월

서울대학교 대학원
전기컴퓨터공학부
정 영 모

정영모의 공학박사 학위논문을 인준함
2019년 8월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Abstract

Image processing methods for light-field based computational imaging

Youngmo Jeong

Department of Electrical Engineering and Computer Science

College of Engineering

Seoul National University

Due to the enormous commercial success of mobile devices that include optics and image sensors, researches are continuing on next-generation cameras that support computational imaging, rather than simple imaging technique. The miniaturization of image sensors and the improvement of image processor performance led to a soft landing of such researches and commercialization. Also, there is a growing demand for imaging systems and image processing supporting them that utilize the latest technologies and support multi-functionality.

In this dissertation, image processing techniques which consider their optical configuration and properties for computational imaging and optical reconstruction methods are studied primarily for light-field based images. A light-field camera is considered as a proper candidate to replace a conventional mobile camera by capturing depth and color information at the same time. However, previous research on light-field cameras tends to be devoted to improving image post-processing technique without considering its optical configuration rigorously. To design the light-field camera and its image processing for optical reconstructions, the composition of the optical elements and system configuration of the light-field camera should be considered and referred.

Considering the above, design methods of the light-field camera are studied,

including accurate F-number matching condition, adaptive depth extraction method based on optical configurations. The proposed design and reconstruction methods utilize several camera parameters. Since the image processing methods in this dissertation comprehensively consider specifications of the system, all information extracted from the image post-processing has real distance units. Therefore the depth-map images and three-dimensional models can be transformed into elemental images for next-generation autostereoscopic display, which can be used for virtual and augmented reality.

Also, more extensive researches are carried out to design and optimize light-field based computational imaging technology using a deep learning software library. Conventional computational imaging techniques are developed in terms of image processing independently or using heuristic cost function for freeform optics. In this dissertation, a technique is proposed to simultaneously optimize optical elements and image reconstruction methods suitable for computational imaging using ray-tracing based technology. Moreover, non-paraxial lights are considered to enhance the quality of reconstruction on the peripheral region of the images. The entire pipeline from input images to reconstructed images is implemented to be fully differentiable for updating variables with stochastic gradient method. The designed device enables extended depth of field through computational imaging and included the design for off-axis input light so that the proposed optimization technique can be used in optical design with larger numerical apertures.

The several image processing methods for light-field based computational imaging technology proposed in this dissertation are expected to be used as future imaging technology and also work with a camera system capable of acquiring information for a next-generation display.

Keywords : Light-field camera, computational imaging, image processing, augmented reality, lens design, freeform optics

Student Number : 2013-20878

Contents

Abstract	i
Contents	iii
List of Tables	v
List of Figures	vi
Image processing methods for light-field based computational imaging	1
Chapter 1 Introduction	1
1.1 Image processing of light-field imaging system and computational imaging	1
1.2 Purpose of this dissertation	7
1.3 Scope and organization	10
Chapter 2 Real-time depth controllable capture and 3D reconstruction with a light-field camera	11
2.1 Introduction	11
2.2 Depth Controllable light-field capture using a light-field camera	13
2.2.1 Accurate F-number matching condition	13
2.2.2 Extended analysis of light-field information on depth controllable light-field camera	22
2.3 Three-dimensional reconstruction algorithm from the captured light-field	25
2.4 Real-time depth controllable light-field capture and reconstruction system	29
2.4.1 Light-field display system and the light-field camera	29
2.4.2 Real-time light-field processing and display results	33
2.5 Conclusion	36
Chapter 3. 360-degree light-field capture and reconstruction with a depth extraction algorithm for light field camera	37
3.1 Introduction	37

3.2 Optical structure of light-field camera and depth extraction method.....	40
3.2.1 Pickup process of light-field camera based on geometry optics.....	40
3.2.2 Depth extraction method with optical flow for light-field camera ..	41
3.3 Simulation and analysis for one-shot 360-degree light-field capturing	53
3.3.1 Analysis for the 360-degree light-field capturing	53
3.3.2 Four-dimensional light-field simulation using sub-image synthesis	58
3.4 Experimental results for the light-field capturing and reconstruction.....	61
3.4.1 Experiments for one-shot 360-degree light-field capturing.....	61
3.4.2 Experiments for optical reconstruction of the light-field.....	63
3.5 Conclusion	65
Chapter 4. Ray-based end-to-end pipeline for optimization of optics and image	
post-processing.....	66
4.1 Introduction.....	66
4.2 Non-paraxial ray tracing simulation for computational imaging for	
optimization pipeline	69
4.2.1 Computation of point spread functions with ray tracing method	69
4.2.2 Simulation for image acquisition at image sensor stage.....	74
4.3 End-to-end optimization pipeline for designing optical element and image	
post-processing	76
4.3.1 Structure of end-to-end optimization pipeline	76
4.3.2 Optimization of freeform lens and image processing for extended	
depth of field.....	81
4.4 Conclusion	87
Chapter 5. Conclusion	89
Bibliography.....	91
Appendix	98

List of Table

Table 2.1 Specifications of the proposed system	34
Table 3.1 Specification for simulation and experiments	52

List of Figures

Figure 2.1 Configuration of LFC focused on infinity and its sub-aperture and corresponding pixels.	14
Figure 2.2 (a) Conventional F-number matching and (b) Accurate F-number matching.....	16
Figure 2.3. Monte-Carlo simulation results of (a) Conventional F-number matching and (b) Accurate F-number matching.	16
Figure 2.4 Configuration of depth controllable LFC.	19
Figure 2.5 Real and virtual image space by depth controllable elemental image pickup.....	21
Figure 2.6 Magnified configuration of LFC focused on infinity (not drawn to scale).	22
Figure 2.7 Captured light-field by proposed elemental image pickup system on light-field plane.	24
Figure 2.8 Elemental image generation algorithm for captured light-field.....	28
Figure 2.9. Experimental setup for depth controllable elemental image pickup: (a) overall proposed setup, (b) LFC and inserted MLA, (c) display lens array.....	30
Figure 2. 10 Captured raw light-field depending on various focal lengths: (a) focus on infinity, (b) focus on 1.2m, and (c) focus on 1m.....	32
Figure 2.11. Outermost perspective views from captured raw light-field.....	33
Figure 2.12 Perspective views of the 3D reconstruction from captured light-field using InIm.	35
Figure 2.13 Synchronized video of reconstruction window and overall proposed system.....	35
Figure 3.1 Geometric structure of LFC.....	41
Figure 3.2 Depth extraction from captured light-field using optical flow in (a) integral imaging and (b) light-field camera.....	43

Figure 3.3 (a) Sub-image extraction from light-field image, (b) optical flow calculation between corresponding points, and (c) schematic diagram to derive equation (3.1).	46
Figure 3.4 (a) Extracted depth of center sub-image with seventh sub-image (among seven sub-images) according to optical flow (b) Maximum optical flow on location of sub-image (among seven sub-images).	50
Figure 3.5 Depth extraction example.	52
Figure 3.7 Blur size depending on depth of images.	55
Figure 3.8 Simulation results for 360-degree light-field capturing and DOLF.....	59
Figure 3.9 The 360-degree light-field capturing setup and experimental results.	62
Figure 3.10 Top view of the holographic display system with anamorphic optical transformation model.	64
Figure 3.11 Optical reconstruction of captured information with holograms.	64
Figure 4.1 Schematic diagram of the imaging system for the proposed computational imaging.	72
Figure 4.2 PSF calculation pipeline. Zernike coefficients are used to encode the target surface and declared as trainable variables.	74
Figure 4.3 Image processing pipeline at the on sensor stage.	75
Figure 4.4 Image post-processing pipeline for calculating losses.	79
Figure 4.5 Data flow of the proposed end-to-end optimization pipeline. Trainable variables are updated by SGD algorithm.	80
Figure 4.8 (upper row) Comparison of images from 2 D between ground truths, reconstruction results with 1 D PSF, reconstruction with optimized PSF and (lower row) PSF representations of 1 D, optimized kernel and 2 D.	85
Figure 4.9 (left) PSNR plot according to the distance of subjects and (right) independently trained kernels for non-paraxial input light.	85
Figure 4.10 PSNR plot according to target depths range	87

Chapter 1 Introduction

1.1 Image processing of light-field imaging system and computational imaging

A light-field is a function that approximates an object, an image, or an arbitrary optical information with a ray on a specific depth plane and expresses the intensity information in positions and angular directions [1]. The light-field converted into digital data is the light sampling method on the intensity and direction of light with pixelated structure and the recording method of the distribution of light on the three-dimensional (3D) space. In general, the intensity of light is represented by R, G, and B channels using representative wavelengths. Therefore, the light-field can be understood as a function of the light's intensity distribution with four inputs, a domain with x, y coordinates in space and the angular θ, ϕ coordinates and three intensity outputs for each color channel. In other words, when the spatial position and the angular direction of a ray are determined, the brightness or color information of the ray can be obtained through the light-field function. Additionally, if information about multiple rays can be defined in a particular range of plane, one can say that the light-field function of a particular depth (z) is known. Such light-field information has similar context to complex information of the wavefront that can deductively track the intensity and direction of light of a specific plane in terms of wave optics. If a wavefront is determined as a complex function, the complex information at a different depth can be calculated by wave-based optical propagation represented by Fourier optics [2]. The advantage of wave-based propagation is that, given the information on optical systems, the propagation of complex information can be defined using deterministic equations. Using such mathematical and physical calculation methods, wave-based optics created a number of research fields including holography, and has been actively

used for designing and implementing diffraction-based optical elements that cannot be easily approximated by ray optics. However, understanding the intensity information of a light on a specific surface based on a wave optics is difficult to use in various scope of dimensions. In particular, the existence of a maximum diffraction angle, which is the limit of the diffraction angle produced by a specific size of slit, gives a huge limitation to such a method of modulating light. For example, for optical elements that provide refraction or diffraction greater than the maximum diffraction angle, it is very difficult to calculate the perfect simulations for wave-based optical analysis. In spite of its precision, many types of optical element design software still adopt a ray tracing based design method on a macroscopic level, and a wave-based analysis is used to cope with the physical phenomena occurring in the diffraction region. The trend in optical design based on ray tracing is due to the development of computing power and yet limited computing memory. To increase simulation correctness, the complex information of light should be sampled at smaller intervals. Although the available memory of workstations has increased exponentially, it is still not enough to design a general optical element such as a lens with only a diffraction-based optical system. The development of distributed computing and parallel computing also facilitates the acceleration of ray tracing based simulation, which is considerably easy to be parallelized, and is another reason why research using light-field information is actively underway. Wave-based optical simulation can take advantage of these new computing methods also, but the use of rays in a light-field domain, where one ray is the minimum atom of the computation, is now a priority.

In general, the light-field information having four input variables can be converted into a two-dimensional distribution information by integrating information into a space or angular direction. Particularly, angular information is often integrated in this manner, expressed as θ, ϕ in this dissertation. If the angular information of light-field at spatial coordinate x, y is summed up, intensity value

can be obtained on the point in the space. The intensity information is a 2D distribution of light that can be obtained through a general imaging system, and this process is generally referred to as "taking a picture". Thus, obtaining light-field information can be thought of as a comprehensive version of the concept of photographing. The obtained light-field has information that each pixel diverges or converges in the direction of the angle, so that the user can restore the image in a space other than the depth at which the photograph is taken. In other words, knowing or recording the light-field information at a certain depth means that 3D information of the object can be analyzed in the plane. By reconstructing images along the depth direction using ray tracing, it can be seen that the focus of a specific object is formed or blurred. This makes it possible to deduce how much a particular object is located in the space away from the surface on which the light-field is recorded. In other words, obtaining the light-field information means that 3D information of the object can be obtained, and that the depth map of the photographed image can be deduced and calculated.

The simplest form of light-field camera (LFC) started with integral photography using a lenslet array, which can capture light-field information on a specific focal plane [3]. Intuitively, the light-field information simultaneously contains spatial information and angular information located in opposite domains in Fourier domain. Therefore, it is necessary to construct a structure capable of simultaneously capturing intensity and angular distribution by the position of light. A lenslet array or a microlens array (MLA) is an optical element in which curved surfaces for small lenses are arranged in a form of array. In this dissertation, the optical element is represented by an MLA and the lenses are arranged in orthogonal composition with widths in the order of tens to hundreds of micrometers. When the focus of the MLA is at infinity with respect to the image sensor plane, each lens collects light at the pixel where its optical axis lies. A bright pixel will be recorded for each position corresponding to the center of each lens,

and it can be said that the intensity information of the collimated light is sampled by the total number of lenses constituting the MLA. Let's assume that a little tilted collimated light enters the MLA. Assuming that each lens has no aberration for a tilted input light, each lens produces a bright spot slightly off center from the pixel where the optical axis lies. Likewise, one lens records one intensity information, but the intensity information is recorded in a different position compared to the on-axis situation. How far the information of the light is recorded from the optical axis of the lens is intuitively determined by how much the incident light is tilted. In other words, the position where the light is recorded in image made by a microlens, which is called as a 'lens image' in this dissertation, is the angular information of the incident ray. As a result, when an object is captured using the MLA, light intensity information in the spatial domain is recorded by the number of lenses, and information of each direction of light is recorded in the angular domain by the resolution of the lens image. If the pixels of the image sensor are finite, the total amount of light-field information that can be captured using one sensor is limited to the number of pixels. Accordingly, the rate at which the four-dimensional (4D) light-field information is sampled is changed according to the configuration of the MLA. Particularly, if the MLA includes one lens, then the imaging system can only record each direction information. Conversely, if one lens is assigned to each pixel of the sensor, the imaging system will be able to record only the brightness of the incident light, not angular information. Therefore, the sampling ratio of the position and angular information is determined by the ratio of the width of the MLA lens covering the sensor surface to the pixel size of the image sensor. The LFC proposed in this dissertation is designed to have a size that one lens is responsible for about 10×10 imaging pixels.

In the past, when photosensitivity films were used as media to record images, light-field information could be recorded using a lenslet array that could cover the size of the film. This imaging method, called integral photography, allowed the

light field to be recorded as much as the size of the film [3]. However, the integral photography based light-field recorder was unable to provide light-field information for space larger than the film size. Also, the newly introduced digital imaging sensor was generally made smaller than the film, and therefore this problem was further exacerbated. To solve the problem, people devise a way to insert an MLA into an imaging system with a typical lens-sensor structure. In this dissertation, the single lens used in the camera is called main lens. In a LFC structure, including the main lens, the MLA is located at a distance of the focal distance of the MLA from the image sensor. The main lens is placed so that the focal plane of the main lens is the MLA plane. To analyze the structure, it is generally assumed that the focal length of the main lens is much longer than the focal length of the MLA. As a result, the light converging to the MLA by the main lens is considered as light coming from infinity for the MLA.

A conventional LFC is an imaging system having an optical structure including the main lens, the MLA and the image sensor as described [4]. However, in recent years, the term of LFC is also an expression that collectively refers to imaging systems capable of acquiring light-field information. Therefore, a camera that can acquire spatial and angular information can be called an LFC. However, the LFC in this dissertation means an imaging unit with one main lens, an MLA and a single image sensor.

Images taken with an LFC have a form that is considerably different from normal photography. In order to convert it into a general photograph, it is necessary to post-process the captured light-field images. As mentioned earlier, transforming an image so that people can recognize it means integrating the information of the angular coordinate according to each point on spatial coordinate. While post-processing the light-field images, different depth images can be reconstructed according to the directions of integration and virtual aperture values or shapes can be simulated by different integration methods. In other words, in

order to use an image captured by an LFC, it is necessary to use an image processing using a computer software. In designing a light field camera, it is indispensable to consider an image processing method. Therefore, designing an LFC starts with defining the optical performance of the imaging system by determining the relationship to the key components of the LFC, and is finalized by developing an image processing method that considers the specifications of the system.

Meanwhile, computational imaging technology including LFCs has attracted much attention because it can provide more functions than the general imaging technology to a camera mounted on a mobile device. In the case of conventional LFCs, additional optical elements such as MLA have to be inserted in the main optical path. However, recent computational imaging techniques does not apply these additional optical elements, or create general imaging results simultaneously with results of computational imaging. Computational imaging techniques are developed to give additional functions besides imaging, including depth extraction, hyperspectral imaging and extended depth of field [5]. Also, designing domain-specific optical elements are also researched, which makes computational imaging exploit tailored optics for its own purposes.

To produce optical elements for specific computational imaging, researchers establish a criterion to see the characteristics of the optical elements easily. Then, an indicator for evaluating the reference value is made, and an optimized optical element having a characteristic satisfying the indicator is used. For example, in the case of a lens for collecting light, a point spread function (PSF), which is the intensity distribution of light when non-tilted collimated light passes through the lens, is considered to represent the lens characteristics. In other words, a desired PSF for the lens is assumed and an optical element is designed to find a lens or a diffractive surface having the PSF to design a lens.

The PSF engineering method is encouraging in that it can make it possible to

design domain-specific optical elements. However, researchers have often used heuristic cost functions as an indicator to evaluate PSF. Specifically, it is expected that the specific PSF gives the best results for the computational imaging using the size, shape or distribution of the PSF. Such an approach, however, sometimes cannot give a sufficient degree of freedom to the optimized result, and the optimized result dwells on a search of local minimum.

To overcome these limitations, research is conducted on optimizing optical elements and image reconstruction methods using a deep learning library at the same time. This technique, called end-to-end optimization, can be implemented using the auto-differentiation function to make all the pipelines differentiable, from the surface profile of the device to the reconstruction of the captured image. By using the stochastic gradient descent (SGD) algorithm, the optical elements and parameters to be used for image reconstruction can be optimized at the same time. However, until now, the end-to-end optimization method has very limited surface curvature and numerical aperture and needs considerably large computational resources since the framework is constructed by wave-based propagation [5]. Also, non-paraxial light is neglected for updating and optimizing a certain surface for imaging and several deconvolution algorithm. Development of end-to-end optimization techniques based on light-filed information with ray-tracing methods is indispensable in order to produce lenses that can be used by modern imaging systems.

1.2 Purpose of this dissertation

This dissertation aims at developments of several image processing techniques for light-field based computational imaging including LFC and end-to-end optimized imaging systems. Existing research on LFCs has a limitation that it does not fully reflect the optical configuration of LFCs. In addition, the definition of the basic optical rule of the LFC-matching F-number has not been rigorously defined.

Furthermore, since the camera information is not reflected when reconstructing the photographed light-field image, the reconstructed image so far is extracted from the light-field information only considering the relative size and depth of the object. In order to overcome these limitations and apply the LFC to practical use and reconstruction, studies on the optical design of the LFC and the image reconstruction method considering it in this dissertation are proposed.

Specifically, an accurate F-number matching condition between the main lens and the MLA which are the two optical elements constituting the LFC, is presented. The sampling rate of the 4D light-field surface is then determined according to the relationship between the lens pitch and the focal length of the MLA and the specifications of the image sensor. The conventional LFC had little analysis of the light field information according to the focus position of the main lens. In this dissertation, an analysis of the change in light-field information with respect to the focus position of the main lens is also performed. The angular information of the light field, which is limited by the focus position of the main lens, is also analyzed. A light-field processing method that considers the information of the optical elements of the LFC and the image sensor enables extraction of the sampled interval of photographed light-field information in real distance units. This helps to acquire 3D information reflecting the actual size of an object photographed by an LFC. Methods for making elemental images, which are core applications that can demonstrate these advantages, reconstruct the 3D model obtained by using next generation display technology which can be used in augmented reality and virtual reality. In particular, integral imaging and holographic display technology, which are autostereoscopic 3D display technologies, are used, and it was confirmed that the model was reconstructed to an actual size object by 3D modeling. New algorithm for depth map extraction is the one of the key applications of LFCs. Like the 3D object reconstruction using the light field camera, the depth map extraction method proposed in this dissertation includes information about the size and depth

of the actual object. Since images taken using an LFC have coincident points with very narrow baselines, the intervals of the corresponding points are measured using an optical flow calculation method with sub-pixel accuracy. The optical flow values of the corresponding points are interpreted as having a certain depth information according to the parameters of the camera. The validity of the simulation and the real object is confirmed by using the LFC and the proposed depth extraction algorithm. The proposed depth extraction algorithm has the advantage of restoring an object to its actual size rather than the existing algorithm, and an application for taking 360-degree information of an object at once is proposed. Using a pair of mirrors, the 360 degree information of the object is captured and reconstructed using the LFC. The captured 3D information is optically reconstructed through the holographic display.

Meanwhile, a ray tracing-based end-to-end optimization pipeline that simultaneously optimizes domain-specific optics and reconstruction algorithms for computational imaging is proposed. Conventional computational imaging methods usually focus on restoring images which have been already captured. Although studies on designing optical elements for specific computational imaging are proposed, it is difficult to prove that the actually designed optical element is the most optimized element for the computational imaging because of heuristic cost functions for PSF engineering method [6]. Therefore, there is a need for end-to-end optimization that optimizes optical elements and image reconstruction methods at the same time. Conventional end-to-end optimization uses a wave-based optical system. However, as mentioned above, it is difficult to design a commercially available numerical aperture lens as a limitation of the maximum diffraction angle. To overcome these limitations, an end-to-end optimization pipeline based on ray tracing is proposed and related problems for implementing are introduced and studied. As a result, a computational imaging system for extended depth of field has been designed and a new reconstruction method has been proposed for

improved extended depth of field technology. The proposed end-to-end optimization pipeline enables the creation of lenses of unrivaled numerical apertures compared to previous studies and realizes extended depth of field of higher performance.

1.3 Scope and organization

The contents of this dissertation are largely divided into three chapters. The first two chapters cover the optical design of the LFC and image post-processing for reconstruction of the actual 3D objects. The proposed formulas and algorithms for two chapters are proposed to reflect the optical characteristics and configuration of the LFC. Also, according to the topic of each chapter, the captured light-field image is reconstructed with real-time integral imaging or reconstructed through a holographic display. In the third chapter, research is conducted on optimizing optical elements and image reconstruction methods for more general computational imaging. Designed computational imaging targets extended depth of field.

The main scope of this dissertation is shown in the following note.

- Depth-controllable LFC and real-time reconstruction
 - Accurate F-number matching condition
 - Depth controllable light field capture and analysis
 - Real-time light-field reconstruction with integral imaging
- 360-degree light-field capture with novel depth extraction algorithm
 - Light-field analysis on object plane from light-field camera
 - Depth extraction algorithm for LFC based on camera parameters
 - 360-degree light-field imaging and optical reconstruction
- Non-paraxial and ray-based end-to-end optimization pipeline for optics and image post-processing
 - Ray tracing simulation for computational imaging
 - Simulation for optimized lens and deconvolution
 - Non-paraxial and ray-based end-to-end optimization pipeline
 - Optimization of lens and image processing for extended depth of field

Chapter 2 Real-time depth controllable capture and 3D reconstruction with a light-field camera

2.1 Introduction

Autostereoscopic 3D display is one of the most remarkable technologies among future display technologies [7-11]. Integral imaging (InIm), one of the most actively studied autostereoscopic display technologies, provides quasi-continuous view points and full-parallax [11-13]. The source images for InIm, called elemental images, can be captured by placing lens array in front of a camera. The elemental images can be also generated with computer system and several methods have been researched [3, 13, 16]. There have been issues on capturing process of elemental image, such as pseudoscopic problem and misalignment of optic axis causing distortion of reconstructed 3D image [11]. Various methods are proposed to solve these issues such as correcting elemental images or improving pickup system [16, 17]. There have been many researches to improve the pickup system of elemental image [18-23]. However, since these studies are based on a conventional InIm pickup scheme, capturing an object through a lens array, only the object located beyond the lens array can be captured. To overcome the fact that the object located in between camera and lens array cannot be taken through InIm pickup system, a research on a method for recording back and front of lens array was progressed [23]. This method, however, uses one lens array, a main lens, and a few of relay lenses to make a 4-f system which accompanies with complicated optical alignment and bulky system. On the other hand, the reconstruction plane of object can be adjusted by using pixel mapping algorithm. However the expressible reconstruction plane is very selective and discrete because the algorithm cannot control the actual focal plane of pickup system [19].

The pickup and reconstruction systems which consist of depth-control lens and lens array for InIm have already been suggested [24, 25]. However, the conventional methods of pickup and reconstruction systems have limitations about their implementations and operation. Specifically, a conventional system of pickup and reconstruction in InIm held out the possibility that a set of light-field can be reconstructed by using a novel optical device which cannot be easily fabricated [24]. In contrast, the proposed study has the advantages of high applicability in terms of recording and broadcasting of 3D information. An MLA which is quite a simple and inexpensive element and provide a real-time pseudoscopic to orthoscopic conversion processed by computer, not a high-priced optical element is used. Also, the real-time pixel mapping algorithm is much more flexible in reconstruction system than a stationary optical system. As a result, the proposed system has a merit of being applicable to any type of InIm by revising the pixel mapping algorithm.

Recently, a handheld LFC which is widely used in capturing 4D light-field has been reported [4]. An LFC consists of a conventional camera and an MLA and the spatio-angular information of 3D object can be captured by its special configuration. The fabrication of the LFC is very simple and it is possible to capture the 3D information like the operation of conventional camera device. In general, the LFC is used for still images, for example digital refocusing or synthetic aperture technique, and hardly used to generate source image for 3D display [4, 26-31]. With the LFC scheme, the overall pickup system has an advantage of being portable to run. Even though the concept of LFC has already been introduced, real-time pickup and display of 3D images are totally new applications of LFC. The captured light-field information was analyzed on light-field plane to show which light-field is recorded on specific part of image sensor. Since the main lens of LFC can image its MLA and image sensor to object plane, the real and virtual objects can be captured simultaneously. Not only that, this lays

the groundwork of its expansion into other elemental image of autostereoscopic display, such as multi-view system and holography, by analyzing a full-parallax information from LFC as we will mention in the following.

In this chapter, a depth controllable InIm pickup method with real-time LFC setup is proposed. The proposed LFC is applied to capture a source of 3D imaging. The pickup plane can be adjusted by controlling the distance between MLA and main lens. Hence, the elemental image for InIm can be captured at various focal planes in purpose. The image of MLA is able to take an elemental image of objects that are located in front and rear of the focal plane at the same time. The captured light-field information is analyzed by geometrical optics and 4D light-field analysis. Extended version of F-number matching condition is proposed. For real-time reconstruction, the high frame rate camera with gigabit Ethernet interface was used. Experiments for the proposed system and analysis are performed to show the validity, and results demonstrated the depth controllable InIm pickup and the real-time 3D image reconstruction.

2.2 Depth Controllable light-field capture using a light-field camera

2.2.1 Accurate F-number matching condition

Figure 2.1 shows the configuration of normal LFC whose focal plane is infinity. To generate a proper elemental image from the captured light-field, one has to thoroughly know which information is included in the captured light-field. In previous researches on the handheld LFC, it is considered that the main lens is located on optical infinity because the each microlens is much smaller than the main lens [4]. Since the chief ray of each microlens is supposed to be parallel to each other, the size of MLA and image sensor are the same. According to the specifications of lenses, however, the main lens and MLA should have different sizes. To fully use the image sensor of LFC in any specifications of lenses, the

image sensor should be slightly larger than the total size of MLA. In this section, the light-field information is analyzed to describe the relationship between two optical elements in detail without that assumption. And the captured light-field information is analyzed on two dimensional (2D) light-field plane which expresses a ray as a point on the coordinate system. Meanwhile, the F-number matching problem in 3D information pickup system adopting a lens array was already well determined, which is also called cross talk elimination [32]. As the studies done in previous researches, an elimination of cross talk can be effectively implemented by using another aperture or a telecentric optical system. However, because this approach may lose the greatest advantage of LFC which is portability, how the size of aperture can be thoroughly determined is analyzed when a 3D information is acquired by handheld LFC.

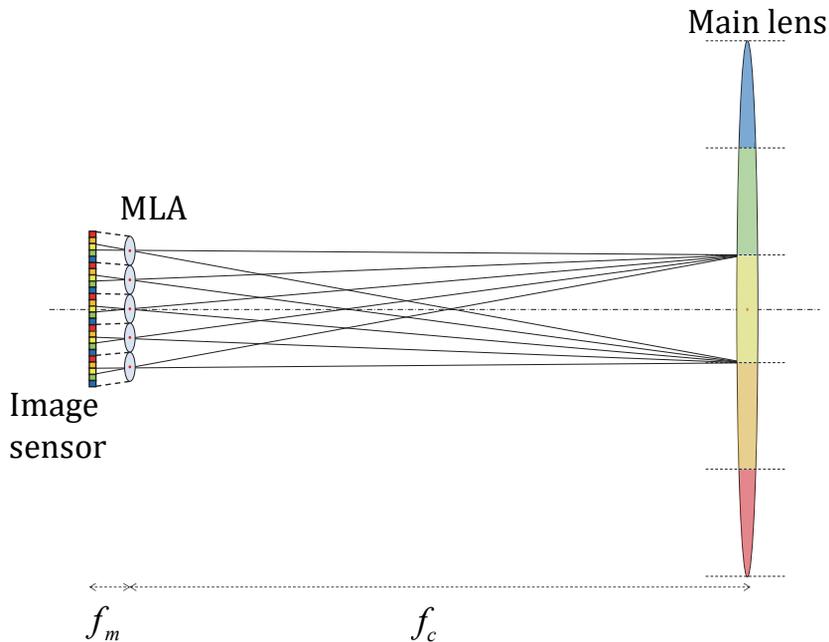


Figure 2.1 Configuration of LFC focused on infinity and its sub-aperture and corresponding pixels.

Firstly, the LFC whose focal plane is on infinity is analyzed. Since the distance between main lens and MLA is the focal length of main lens, only collimated light from infinity is focused on MLA plane. Therefore captured light-field is intermediately focused all over the depth. A bundle of light through a specific region of main lens reaches at the pixels on the same position in local images made by microlenses. Specifically, the aperture of main lens is divided into equal size of sub-aperture which is as many as the number of pixels for one microlens. For capturing the fine light-field data without overlapping or lacking, there is an optical condition called F-number matching. The conventional F-number condition is generally explained as a match of F-numbers of MLA and main lens. Under the accurate version of the F-number matching condition, sub-apertures do not overlap with each other as shown in Figure 2.1. The ray paths through the center sub-aperture are only expressed in Figure 2.1. The accurate version of F-number matching condition can be determined as follows.

The diameter of main lens D , which is usually adjusted by aperture stop, is determined by

$$D = \delta_m \cdot \frac{f_c + f_m}{f_m}. \quad (2.1)$$

The F-number of photographic lens is shown as Equation (1.2):

$$F_c = \frac{f_c}{D}. \quad (2.2)$$

With the F-number of Equation (2.2), LFC not only captures the proper light-field but also maximizes the usage of pixel of image sensor. The number of pixels n_m for each lens image by single microlens is calculated by Equation (2.3), and all lens images have the same number n_m :

$$n_m = \frac{f_m}{\delta_p \cdot F_c}. \quad (2.3)$$

When the size of $\delta_p \cdot n_m$ is slightly larger than the lens pitch of the MLA, δ_m , a set of pixels located on the same position in each image by a microlens record the rays from the same section on main lens. One can find that every center pixel in lens image stores the light-field from the third part of main lens. In comparison with conventional F-number matching condition, the proposed F-number matching condition requires smaller F-number of main lens.

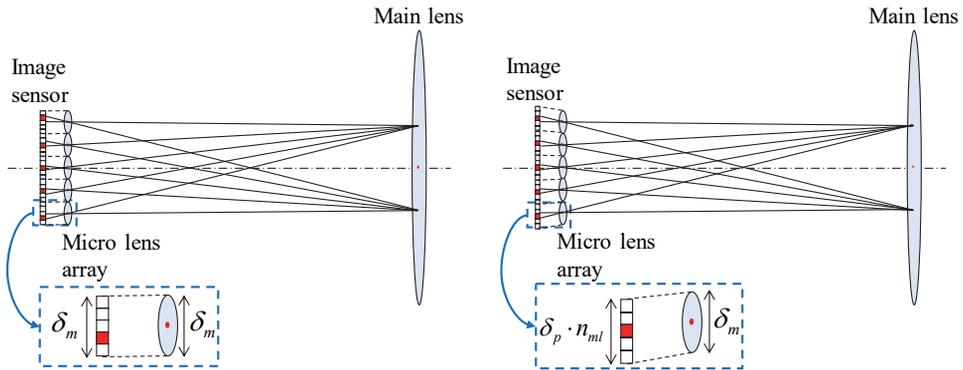


Figure 2.2 (a) Conventional F-number matching and (b) Accurate F-number matching.

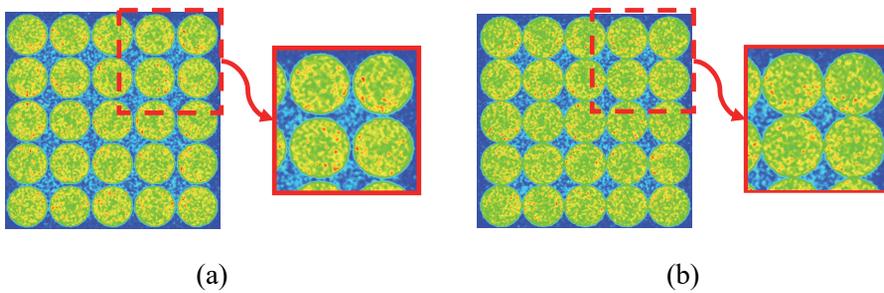


Figure 2.3. Monte-Carlo simulation results of (a) Conventional F-number matching and (b) Accurate F-number matching.

Through the Monte-Carlo based optical simulator, it is confirmed that the proposed accurate F-number matching method delivers higher density of light energy to the image sensor surface. Figures 2.2(a) and (b) show the F-number matching in conventional LFCs and the proposed method. As shown in Figure 2.2(a), using the conventional F-match method does not correspond with the assumption that light passing through microlenses has different optical axis with main lens. It cannot be assumed that the relative positions of images by MLA have passed through the same sub-aperture of the main lens, and as a result, the correct light-field cannot be recorded. Figure 2.2(b) shows the proposed F-number matching method, which shows that the size of the image sensor used is larger than that of the MLA ($\delta_p \cdot n_m > \delta_m$). Therefore it is confirmed that light passing through the same portion of the main lens is recorded at the same position on the image sensor. Using the Lighttools simulator, what image is captured according to the newly defined F value is simulated. An MLA having a circular aperture is used, and the diameter of the main lens is different according to the conventional case and the proposed case. Figure 2.3 shows the simulation results. The left image in Figure 2.3 shows the case where the F value is 4, which is matched by the conventional method. In the right image, the F value is matched through the proposed method when the F value is 3.8. As shown in Figure 2.3, in the proposed method, the empty space between each lens image is reduced, and it is confirmed that the image sensor can be used efficiently by 7%.

Figure 2.4 shows the configuration of depth controllable LFC. The image sensor, MLA, and a main lens compose the LFC. An LFC can capture a light-field information of given scene using a main lens and an MLA on its image sensor. Each ray passing through the main lens and MLA is stored in the different pixels according to the direction of rays [4]. The MLA and sensor are imaged by main lens to subjects. The images of MLA and sensor, which are shown in the right side

of Figure 2.4, are located on the focal plane of main lens. According to the position of the focal plane, the floated MLA and sensor work as an InIm pickup system which is usually composed of large scale lens array and pickup device. Therefore, depth controllable LFC can be used as a pickup system of InIm whose reconstruction plane can change freely. Though the sizes of sensor and MLA are generally described to be the same, the size of image sensor has to be larger than that of MLA as shown in Figure 2.4. If there is no assumption that the focal length of main lens is much longer than that of MLA, the area on sensor made by MLA becomes larger than the size of MLA, as will be analyzed in next chapter.

To control the reconstruction depth of light-field, one can move the focal plane of main lens by controlling the distance between main lens and MLA. It is adjusted by focusing ring of main lens which usually comprises of a group of lenses to control the distance between two optical elements. Figure 2.4 shows that main lens focuses on specific focal plane where a is the distance between the MLA and main lens and b is the distance between the main lens and focal plane which is calculated by lens maker's formula. It is supposed that the focal length of MLA and main lens are f_m and f_c accordingly. The LFC having a focal plane on finite distance makes the distance between the main lens and MLA be the focal length of main lens. These floated MLA and sensor capture the light-field of given scene.

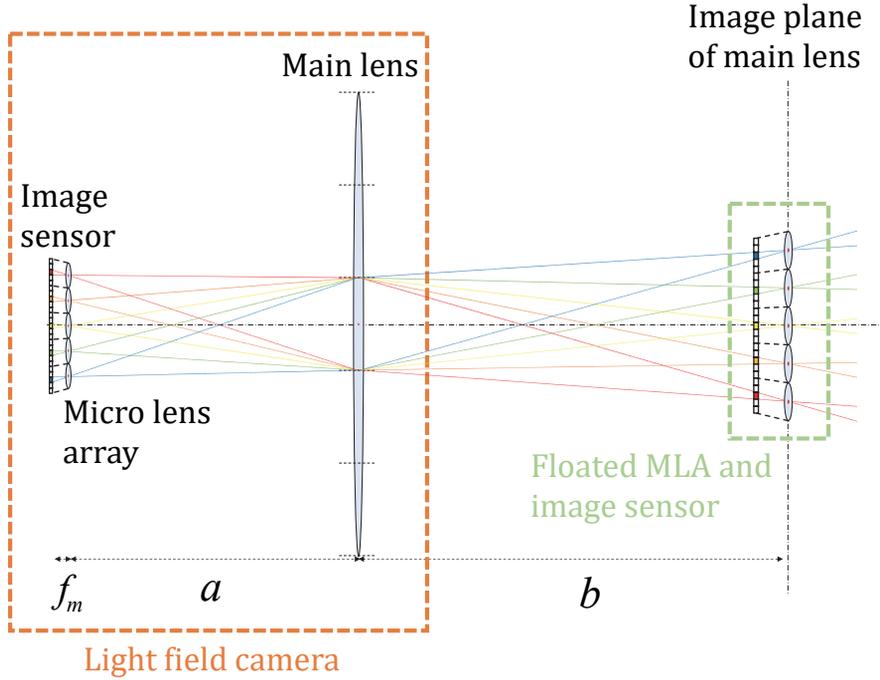


Figure 2.4 Configuration of depth controllable LFC.

Depending on the distance between the main lens and the MLA, the position and size of the floated sensor and MLA can be obtained by using the geometrical optics analysis. The specifications of sensor and MLA for pickup are determined according to the position of focal plane. Specifically, when the main lens of LFC is composed of only one lens, the pixel pitch δ_{pp} and the lens pitch δ_{lp} of pickup system shown in Figure 2.4 are calculated as

$$\delta_{lp} = \frac{b}{a} \cdot \delta_m, \quad (2.4)$$

$$\delta_{pp} = \frac{b'}{a + f_m} \cdot \delta_p, \quad (2.5)$$

where δ_p , and δ_m are pixel pitch of image sensor and lens pitch of MLA,

respectively, and b' means the distance between main lens and image of sensor which is calculated by lens maker's formula. The focal length of pickup lens array shown in Equation (2.5) can be determined as a gap between the image of sensor and image of MLA:

$$f_l = b - b'. \quad (2.6)$$

Figure 2.5 shows the proposed elemental image pickup system which is floated by the main lens of LFC. Since the image sensor is located on the left side of MLA, the pickup direction is represented as rightwards arrow. The overall size of floated image sensor is smaller than floated MLA. The microlens and image sensor capture a given object by leaner way to optical axis of system. In terms of elemental image of focused mode InIm, which is employed in the proposed reconstruction system, the optical axes of lens array have to be parallel to each other [17]. In the proposed elemental image pickup system, however, each floated microlens has different optical axis. Because of that, a part of the captured light-field may have large angular information which cannot be provided by the InIm system. To properly use the LFC as a pickup method of elemental image, the captured light-field data are modulated according to the specifications of reconstruction system which is analyzed in following chapters. However, if the F-number of MLA is large enough to utilize a paraxial approximation for incident light, an image of MLA and sensor can be considered as pickup system of elemental image for focused mode InIm.

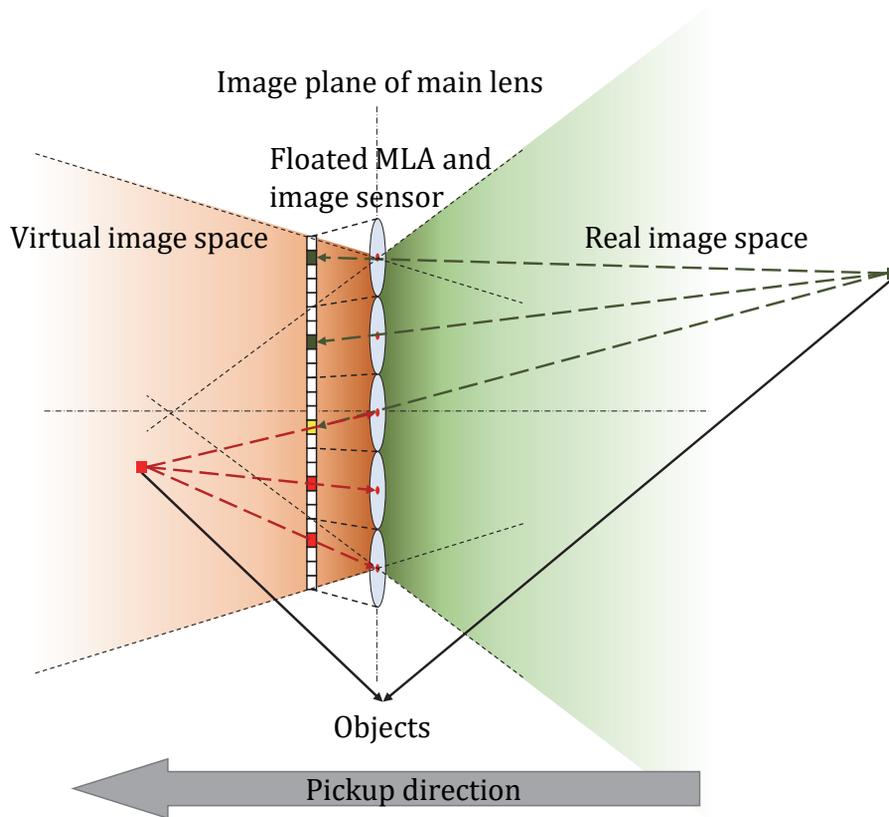


Figure 2.5 Real and virtual image space by depth controllable elemental image pickup.

Unlike the conventional elemental image pickup, the light-field information of the objects located on the front and the rear of the pickup system can be obtained without additional optical devices at the same time as shown in Figure 2.5. The magnified sensor and MLA make virtual image space and real image space. Each image space is spatially distributed within the depth of field (DOF) of main lens. In focused mode of InIm, the plane of lens array for pickup is reconstructed at the plane of lens array for display. Therefore, the plane of display lens array becomes the focal plane of LFC. The objects of real image space and virtual image space are accordingly reconstructed in front and back of display lens array under the pseudoscopic to orthoscopic conversion [16, 31].

2.2.2 Extended analysis of light-field information on depth controllable light-field camera

Figure 2.6 shows the detailed configuration of LFC. By analyzing capturing parameters, it is possible to estimate the viewing parameters of reconstructed 3D image. The total angular information Ω , is approximately calculated by the reciprocal value of F-number. The total spatial information S , which means the overall size of image sensor, is calculated by $\delta_p \cdot n_m \cdot N$ where N is total number of microlenses, and the number of pixels n_m for one microlens is expressed as five pixels in Figure 2.6. The ω means the tilt angle of a chief ray of light-field from center of main lens to each microlens. Since the image sensor is larger than MLA for accurate F-number matching, the center pixels on each lens image have different angle of chief rays. The tilted angles of each chief rays are calculated as Equation (2.6) where the position of each MLA from center, q , is integer from 1 to N , as follows:

$$\omega_{\pm q} = \tan^{-1} \left(q \cdot \frac{\delta_p \cdot n_m}{f_c + f_m} \right). \quad (2.7)$$

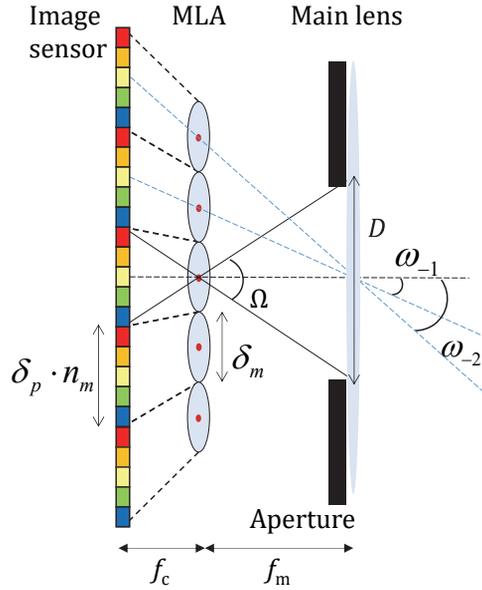


Figure 2.6 Magnified configuration of LFC focused on infinity (not drawn to scale).

In order to reconstruct a captured light-field through a 3D display, it is useful to analyze captured light-field on light-field plane. A 2D light-field can be represented as a point (x, θ) in the light-field plane. In Figure 2.7, x -axis and θ -axis represent the pixel location and the incident angle of light-field on image sensor plane, respectively. The slope of parallelogram is not depicted exactly in accordance with Figure 2.6. The shaded area on light-field plane is the light-field information of given scene when the distance between the main lens and MLA is the focal length of main lens. Each shaded dot means the equally shaded pixels of image sensor as indicated in Figures 2.6 and 2.7. The total size of angular information Ω is determined by the F-number of the main lens as mentioned earlier, and S means the overall size of image sensor. The number of sampling points along the x -axis and θ -axis is determined by the N and n_m . The distances between spatial and angular sampling points are determined by the space between adjacent images made by the microlens, $\delta_p \cdot n_m$, and ω_{+1} calculated by Equation (2.7). Because the center of each microlens and its corresponding part of image sensor have different incident angles, the bundle of light-field recorded on same location on each elemental image represents a parallelogram on light-field plane, not a rectangular. The overall slope of the parallelogram of captured light-field is calculated as $-1/(f_c + f_m)$ as shown in Figure 2.7. It is interesting to note that the amount of angular information of each microlens is supposed to be the same. To be more accurate, the angular information of each microlens is reduced gradually with the distance from the center of image.

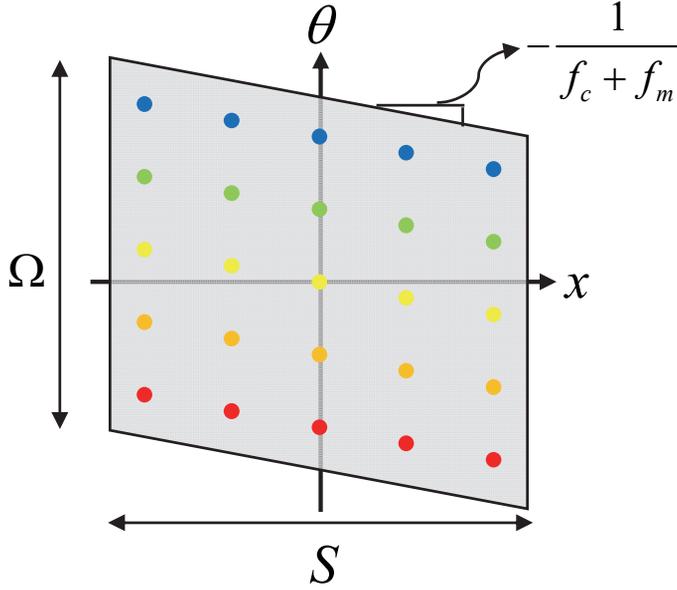


Figure 2.7 Captured light-field by proposed elemental image pickup system on light-field plane.

To implement the depth controllable InIm pickup, the distance between main lens and MLA becomes larger than the focal length of main lens. Consequently, the focal plane of main lens moves closer to main lens. The incidence angle of light-field decreases according to the distance from the sensor, which reduces the size of the image sensor on each of the MLA. This phenomenon creates divided areas on sensor that cannot record a light-field, and hence, wastes image sensor. In other words, the captured light-field is fragmented and it corresponds to the lack of light-field information. To solve this problem, the aperture should be more open according to the distance from the main lens and the image sensor. Specifically, it is necessary to introduce a new F-number matching according to the focal plane, where Equations (2.6) and (2.7) can be used with plugging in a instead of f_c . However, because the value of a does not change dramatically by moving the focal plane from f_c generally, the following analysis was carried out without consideration of these effects.

According to the location of focal plane of main lens, the lateral sizes of two optical elements are enlarged by a factor of magnification, which is calculated as the ratio between the distance from main lens to focal plane and MLA, b/a shown in Figure 2.4. Since the focal length of main length is generally much larger than that of MLA, the magnification values of two optical elements are almost same. With this assumption, the distance between the enlarged MLA and image sensor is $(b/a)^2$ times larger than the focal length of MLA. By considering magnification, the actual specifications of pickup lens array and image sensor can be determined. Given specifications of lens array, image sensor, and display system, the viewing parameters of reconstructed 3D images can be calculated, such as lateral resolution, depth resolution, and viewing angle [33, 34].

2.3 Three-dimensional reconstruction algorithm from the captured light-field

Captured light-field information is processed with pixel mapping algorithm and is transformed into proper elemental image [16, 19]. Even though the captured light-field already has the almost same form of elemental image, the light-field has to be resampled considering specifications of display panel and display lens array (DLA) and reassigned by pseudoscopic to orthoscopic conversion. The elemental image is reconstructed as 3D image at display stage using focused mode InIm method. In terms of light-field analysis, an InIm system in focused mode completely reconstructs rectangular shape of light-field on light-field plane [35]. As shown in Figure 2.7, however, the captured light-field of LFC has a slanted angle on light-field plane. To avoid this difference, the captured light-field has to be sampled in rectangles on light-field plane according to the specifications of reconstruction system. Each elemental image captured by one microlens is assigned to each lens of lens array. As a result, the pixel algorithm in proposed configuration has following procedures: determining the resolution of each elemental image from

given display and camera specification, sampling predetermined number of pixels and location in elemental image according to the F-number matching condition, and mapping each elemental image by rotating it by 180 degrees. From previous research, it is verified that this algorithm can be implemented as real-time processing except resizing process [17, 27].

Figure 2.8 shows the overall algorithm to make elemental image for InIm. Firstly, reconstructed light-field should have same angular information as captured information. In general, an MLA with low F-number is preferred to record large angular information. Therefore the angular information of captured light-field has to be limited according to the maximum angular information which can be reconstructed by DLA. The details of viewing characteristics in InIm were already researched [11, 34]. The expressible pixels have to be selected within the area of light-field. In terms of 2D elemental image, this process is identical to crop and attach the each MLA image of raw light-field as shown in Figures 2.8(a) and (b). Specifically, the size of DLA image is proportional to the ratio of F-number of MLA to F-number of DLA. If the F-number of DLA is two times larger than MLA, for example, the radius of cropped image is half of that of MLA image. The centers of DLA images, which are shown as center points of dotted circles in Figure 2.8(a), are different from each lens image. Since the optical axis of each MLA slants to optical axis of camera, the center of cropped image should be shifted from center of lens image to edge of elemental image. The center deviation of each axis counted by number of pixels is calculated as

$$d_q = q \cdot \frac{n_m \cdot f_m}{f_c + f_m}. \quad (2.8)$$

Meanwhile, the angular information of captured light-field becomes insufficient when the F-number of MLA is larger than that of DLA. Therefore, the final elemental image has lots of empty pixels. If one uses full of lens image for

elemental image to avoid this problem, every light-field has a distortion in its reconstruction direction. As a result, the 3D image space shrinks along the reconstruction depth. In this case, the shrinkage degree of reconstruction depth is also proportional to ratio of F-number of MLA to F-number of DLA.

Second, there is a pseudoscopic problem in captured light-field. A pseudoscopic problem which commonly occurs in elemental image pickup is triggered by different directions of pickup and display [31]. Under the pseudoscopic problem, InIm suffers from reversed depth expression. It is said that pseudoscopic problem in focused mode integral image can be solved by rotating each elemental image by 180 degrees which is shown in Figure 2.8(c) [17, 31]. By using pseudoscopic to orthoscopic conversion (PO conversion) algorithm, the object in virtual image space shown in Figure 2.5 is reconstructed in back of display lens array. On the other hand, the object in real image space is reconstructed in front of display lens array.

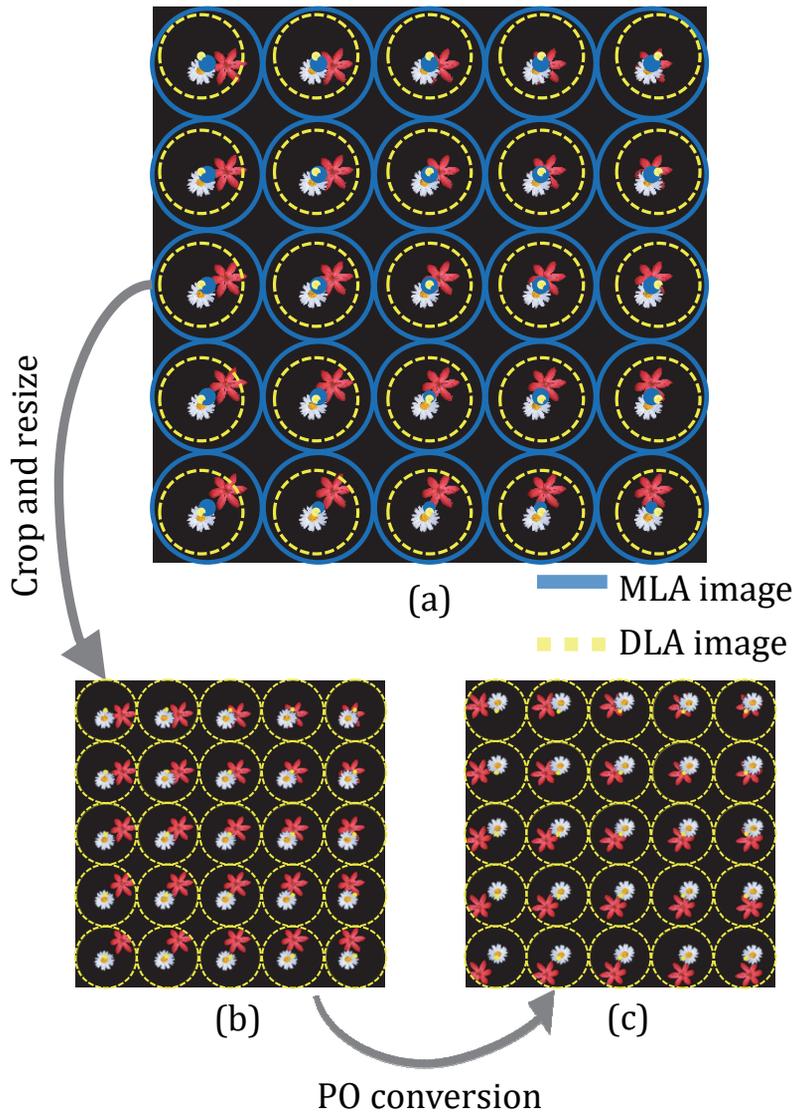


Figure 2.8 Elemental image generation algorithm for captured light-field: (a) MLA image (solid line) and DLA image (dotted line) on raw light-field data, (b) cropped elemental image within the angular range of DLA, and (c) orthoscopic elemental image after PS conversion.

2.4 Real-time depth controllable light-field capture and reconstruction system

2.4.1 Light-field display system and the light-field camera

Figure 2.9(a) shows the experimental setup for the proposed system. The LFC was implemented by locating MLA on image sensor. The MLA and CCD for image sensor are shown in Figure 2.9 (b). Figure 2.9(c) shows the lens array for 3D display stage.

The distance between MLA and CCD sensor is set as the focal length of MLA to record the sharpest light-field. The distance between two optical elements is adjusted by using collimated light. The MLA is fixed with acrylic structure at the location which makes the sharpest recording of collimated light source. The F-number and focal plane of main lens are adjusted through aperture ring and focus ring at main lens, respectively. A light-field captured by high frame rate CCD (Allied Vision Tech Prosilica GX2300C) transfers to computer and is processed by real-time pixel mapping algorithm made by OpenCV without GPU processing. For InIm, high resolution liquid crystal display (IBM T221, 22 inch, 3840×2400) is used as display panel, and 1 mm lens array with focal length of 3.3 mm is used as display lens array. MLA which has lens pitch of 250 μm and focal length of 1mm is used. The number of available microlenses is 43×36 . The focal length of main lens is fixed as 30mm. From the accurate F-number matching condition, the F-number of main lens should be 3.87 by Equation (2.2). Because our main lens (Avenir TV Zoom lens) cannot provide continuous aperture ring, the system used the nearest F-number which was 4.

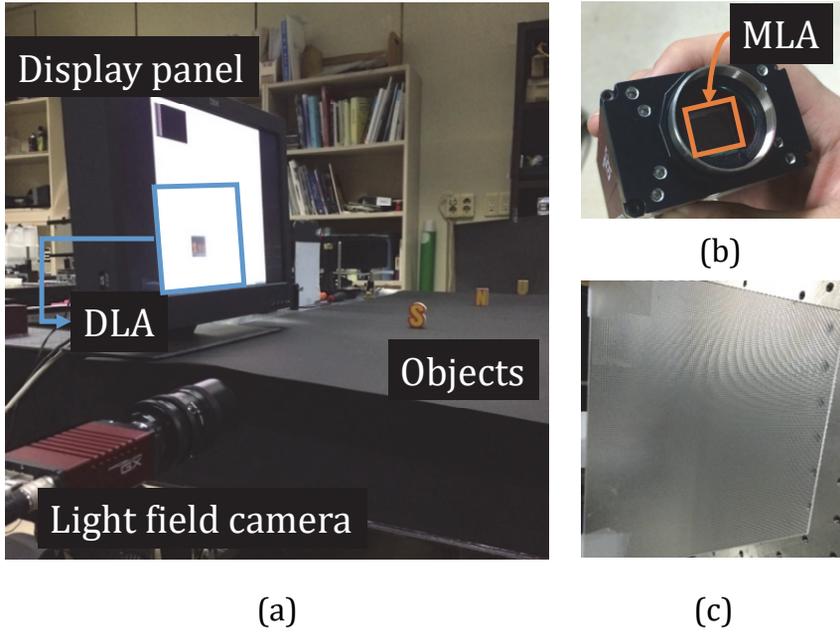


Figure 2.9. Experimental setup for depth controllable elemental image pickup: (a) overall proposed setup, (b) LFC and inserted MLA, (c) display lens array.

To verify our system, three objects of alphabets ('S,' 'N,' and 'U') which are located at different distances (1 m, 1.2 m, and 1.4 m from the system, respectively) are located. The height of toys is around 8 cm. According to the location of focal plane, the actual specifications of pickup lens array and display are determined. When the focal plane is on 'U' where it is located on 1.2 m from main lens, the total size of available area of MLA is $10.75 \text{ mm} \times 9 \text{ mm}$. The actual pickup lens array has lens pitch of 10 mm and total size of $430 \text{ mm} \times 360 \text{ mm}$. The focal length of enlarged MLA is around 65cm using Equation (2.6). Since the active area of image sensor covered by MLA is 2021×1692 and pixel pitch of $5.5 \mu\text{m}$, the actual image sensor on 1.2 m becomes 2021×1692 resolution with pixel pitch of 0.1 mm which is calculated by Equations (2.1) and (2.2).

Figure 2.10 shows captured light-field according to focal plane (infinity, 1.2 m,

and 1 m, respectively) of main lens. For example, Figure 2.10(b) shows captured light-field focused on ‘U’ using our proposed pickup system. As shown in magnified image of Figure 2.10(b), the object is recorded on 8 microlenses in vertical direction because the size of floated microlens is around 1 cm. Based on the focal plane of main lens, objects beyond the focal plane are captured as real images and objects in front of focal plane are captured as virtual images as shown in Figures 2.10(a) and (c), respectively. Magnified images of Figure 2.10 demonstrate that elemental images whose objects in back and front of floated MLA have different record directions from the center of each elemental image. A captured light-field has a pseudoscopic problem because pickup and display directions of light-field are different. Figures 2.10(b) and 2.10(c) show captured light-field focused on ‘U,’ and ‘S’. They show that the position of lens array can be changed by adjusting focal plane of main lens. The captured raw light-field according to the focal plane of LFC is recorded in real-time. Figure 2.11 shows the outermost perspective views from the captured light-field. There is a vignetting problem at periphery part of light-field, which is triggered by lens limitations.

According to given display lens array and display panel, the captured light-field is converted to proper elemental image. Because the pixel pitch of display panel is $124.5\mu\text{m}$ and that of display lens array is 1mm, each elemental image becomes to have approximately 8×8 pixels. Each elemental image is extracted from each circle image in light-field which is 47×47 pixels. Only 8 pixels per one elemental lens could be achieved because the specifications of display panel and lens array limit the viewing characteristics of reconstructed 3D images. If one had used a bigger pitch of lens array in display stage, higher image resolution could have been achieved along with changed viewing characteristics. Meanwhile, if not limited to only InIm, autostereoscopic displays such as multi-view display, which have much higher resolution, can be made without wasting any light-field information.

Because the F-number of display lens array is 3.3, captured light-field could not contain the angular information which can be provided by display lens array. Even though one can use the all of angular information of captured light-field, the reconstructed 3D image cannot help having depth distortion. The reconstructed 3D image has about 83% smaller depth range compared to desirable depth expression as mentioned in the previous section.

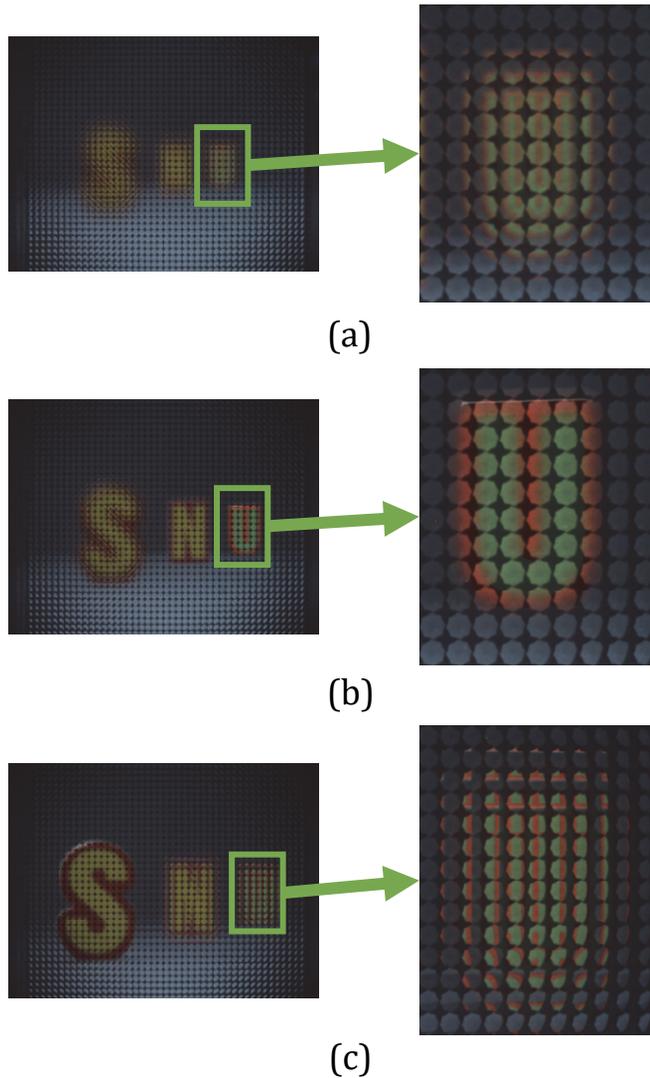


Figure 2. 10 Captured raw light-field depending on various focal lengths: (a) focus on infinity, (b) focus on 1.2m, and (c) focus on 1m.

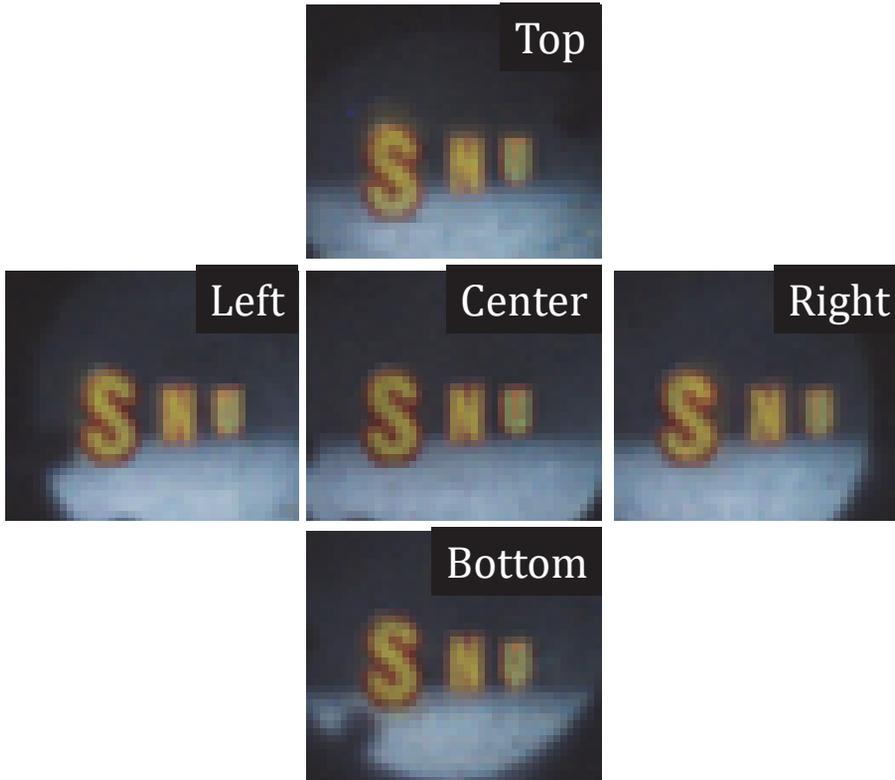


Figure 2.11. Outermost perspective views from captured raw light-field.

2.4.2 Real-time light-field processing and display results

Figure 2.12 shows reconstructed 3D images which have full parallax. Overall proposed system and reconstruction window are presented in Figure 2.13. Figure 2.13 shows synchronized video of reconstruction window and overall proposed system. The reconstruction window shows the 3D reconstructed objects in real-time and overall proposed system is shown in box with solid line. In the Figure 2.13 one can see that the magnification of captured light-field is changing according to focal plane, even though one tries to control the focal plane. The phenomenon which is called focus breathing, is unwanted effect that magnification of lens changes according to its focal plane. Latest commercialized lenses are designed to lessen this problem by controlling their internal lenses, not outermost lens. The detailed specifications of all optical elements in capture and display

systems are shown in Table 2.1. The light-field information is transmitted to computer at 32 frames per second (fps) and it is reconstructed at 16 fps with pixel mapping algorithm without GPU processing in real-time. Though the proposed pixel mapping algorithm is already operated in real-time without delicate method such as GPU processing, additional methods to boost the speed of image processing have to be considered for better broadcasting environment. By using parallel processing, the frame rate will be higher than the proposed system [36, 37].

Table 2.1 Specifications of the proposed system

System elements	Specifications	Value
MLA	Lens pitch	250 μm
	Focal length	1 mm
Display lens array	Lens pitch	1 mm
	Focal length	3.3 mm
Main lens	Focal length	12.5 – 75 mm
	F-number	1.8 – 16
CCD	Pixel pitch	5.5 μm
	Resolution (pixels)	2336 \times 1752
Display panel	Frame rate	32 Hz
	Pixel pitch	124.5 μm
	Resolution (pixels)	3840 \times 2400
Captured light-field	Total resolution (pixels)	2021 \times 1692
	Number of microlenses	43 \times 36
Elemental images	Total resolution (pixels)	344 \times 288
	Size of elemental image (pixels)	8 \times 8

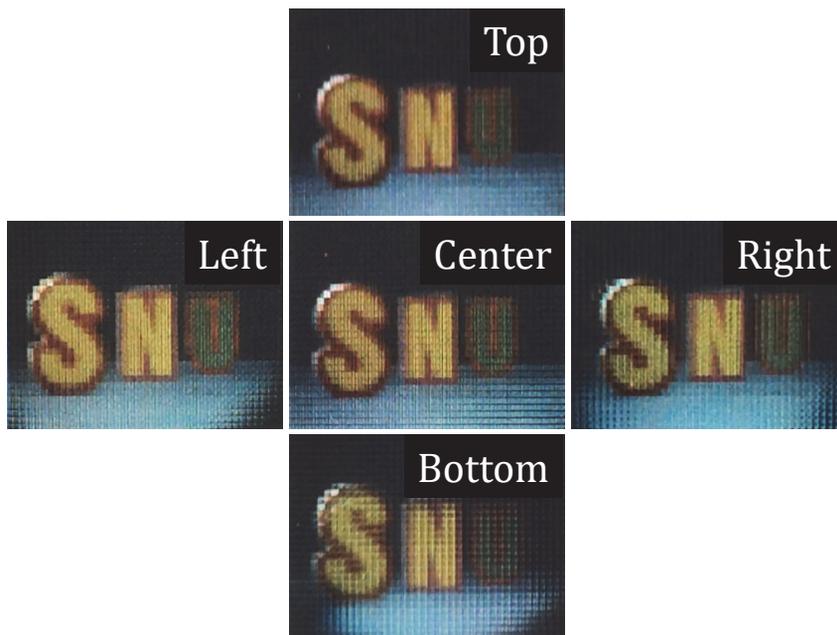


Figure 2.12 Perspective views of the 3D reconstruction from captured light-field using InIm.

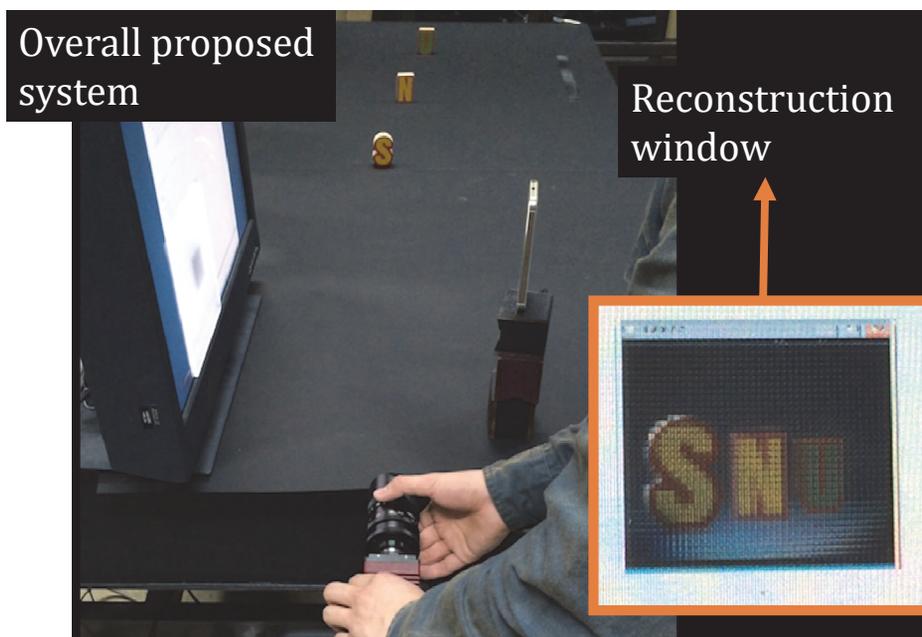


Figure 2.13 Example frame of synchronized video of reconstruction window and overall proposed system.

2.5 Conclusion

In this chapter, a real-time depth controllable InIm pickup with LFC setup which can provide real-time pickup and 3D reconstruction using InIm system is proposed. The LFC is implemented by MLA and high-frame rate CCD which can provide 32 FPS. The MLA is fixed on image sensor by acrylic structure. For F-number matching at LFC, accurate F-number matching condition is suggested. The proposed LFC can capture 2021×1692 resolution which contains 43×36 lens array. For real-time pickup and 3D reconstruction, the pixel mapping algorithm which considers the f-number matching condition at display stage and pseudoscopic problem is used. By using high resolution display panel and lens array, 3D images are reconstructed in real-time. The reconstructed 3D image has full-parallax and is updated at 16 fps. The proposed pickup and 3D display system can be applied to various research fields and broadcasting area based on autostereoscopy.

Chapter 3. 360-degree light-field capture and reconstruction with a depth extraction algorithm for light field camera

3.1 Introduction

With the explosive interest in implementing and improving augmented reality (AR) and virtual reality (VR) systems, there is an increasing demand to acquire information on real 3D objects to produce contents for realistic displays. Efforts to develop core applications of VR and AR displays such as head mounted display (HMD) and head-up display (HUD) are widely researched [38–42]. In order to display natural real 3D scene, obtaining 3D information of the object is necessary. There are several different ways to acquire 3D information, such as structured light, time-of-flight, and light-field [4, 43–46]. Among them, the LFC is attracting much attention as a next generation camera which can record depth and color information of object while it is similar to a conventional camera in appearance and has an advantage in portability [4]. An LFC records 4D light-field (two-dimensional spatial information and two-dimensional angular information) of given scene in forms of two-dimensional (2D) image. Since it was proposed as a device for taking plenoptic function, there have been a large number of studies on ways to handle and reconstruct the captured light-field [27, 28 30, 47, 48]. They also explored how to get more precise information from an LFC and what optical information an LFC gets exactly—in other words, where a ray starts and where it heads to [47, 49, 50].

According to specifications of optical element in LFC, such as sensor size, pixel pitch, lens pitch, f-number and position of focal plane, each pixel of LFC is magnified to a certain position on space and records a bundle of rays having a specific range of angle of incidence. This means that light-field information is sampled differently to captured image depending on many input parameters of LFC.

Therefore, to transform captured image to exact light-field, it is necessary to define which light-field the image sensor samples according to the details of the LFC. One research on the LFC has been conducted on how much information the captured 2D image has in the 4D light-field domain [50]. By using this study, an image taken by an LFC can be easily processed to light-field information. On the other hand, to reconstruct a 3D object from the recorded light-field, depth information should be extracted from the light-field data. Methods for extracting the depth by using InIm, which is an optical system similar to LFC, have been proposed [51, 52]. These methods use stereo matching with pixel-unit accuracy. Therefore, it is impossible to extract continuous depth information and cannot reconstruct accurate 3D volume. There is also a study to extract depth map with sub-pixel accuracy in images taken with an LFC [53, 54]. However, although this method has a high accuracy, there is a limitation that the extracted depth is a relative value between pixels, not absolute value, so that additional calibration is required to reconstruct as 3D model.

In this chapter, depth extraction method based on optical flow for light-field camera (DOLF) is proposed. Optical flow calculation is used to analyze light-field information for several applications such as visual odometry and reconstruction of transparent flow surface [55, 56]. In these studies, the optical flow is calculated between images made by each lens of lens array. There is also a study using optical flow for depth information in InIm, but it has not yet been used to extract depth information by comparing sub-images taken by an LFC [57]. Paying attention to the difference between InIm and LFC systems, the two optical systems to determine the relationship about the optical flow value of the sub-images and the actual depth of the object are analyzed geometrically. The depth information of center sub-image is calculated as absolute value of length under given conditions of the LFC. To provide stability to the algorithm, maximum optical flow value where corresponding points can be located on physically allowed area is determined

according to relative position difference between sub-images. By using this algorithm, one can extract actual depth information from light-field image taken by LFC under given condition of the LFC. Consequently, the optical flow method suitable for LFC images is explained.

A 3D model of the real object can be reconstructed if there are multiple light-field images captured from different location of LFCs. Each light-field image taken by LFC becomes a perspective light-field image from the object, which composes total light-field of the object. If the location and orientation of the LFCs for perspectives are given, the 360-degree light-field information of the object can be determined. To reconstruct the object by using optical or computational system, the depth and color information should be extracted from the perspective light-field. From the captured light-field images, the DOLF can be used to extract depth information and digital refocusing technique can be utilized to calculate clear color images. By binding overall perspective information of depth map and color image, 3D structure of the object can be reconstructed. Therefore, an optical system that acquires 360-degree light-field information of an object by using an LFC in one-shot is also proposed. Putting several depth cameras in multiple locations or using mechanical apparatus to control orientation of subjects is a conventional way to achieve 360-degree depth profile [58, 59]. However, the systems have limitations on both cost and size or the output results are a long way from the high-quality images. To overcome these limitations, a compact and inexpensive 360-degree light-field capturing system is proposed in this chapter. This system can be used as an alternative of 3D scanner. The capturing system allows capturing not only depth and color information of front view but also that of back and side views. The proposed system contains two plane mirrors and one LFC. Two mirrors and an object are placed within the viewing angle of the LFC to obtain 4D light-field information of the object in a single shot. The locations and orientations of two mirrors are analyzed to optimize 360-degree light-field recording and controlled

minutely to determine the angle of pickup direction. Experiments and simulation results are presented to support feasibility of DOLF and the capturing system. 360-degree light-field information is captured in one-shot and optically reconstructed as the 3D object. Holographic display is used to reconstruct the wavefront of the captured objects. Experimental results prove the validity of the proposed pickup method with LFC and DOLF.

3.2 Optical structure of light-field camera and depth extraction method

3.2.1 Pickup process of light-field camera based on geometry optics

Figure 3.1 shows the geometrical structure of an LFC. In this dissertation the standard plenoptic camera is used as the structure of LFC. When the distance between the MLA and main lens is a , the distance from the main lens to the focal plane is represented by b_1 where the focal plane of LFC lies at a specific location, not infinity. The microlens diameter is l_p and the focal length of the MLA is f_m . The light-field information of the object is recorded by the image sensor and MLA magnified by the main lens. Since the pickup unit is floated on the specific position, it has an advantage that all objects in front and rear of focal plane can be photographed. According to the modified f-number matching condition, the overall sizes of the image sensor and the MLA are different, and the distance between two optical components is as much as the focal length of the MLA. Because each element is located at a different distance from the main lens, the location and magnification of two elements differ from each other. By using simple geometrical optics, one can determine the magnification value for MLA, and from this, b_1 is determined with the lens formula. The distance between main lens and magnified sensor is $b_2=(f^{-1}-(a+f_m)^{-1})^{-1}$. Detailed optical properties of the pickup unit depending on specifications and focal plane of LFC have been analyzed [50].

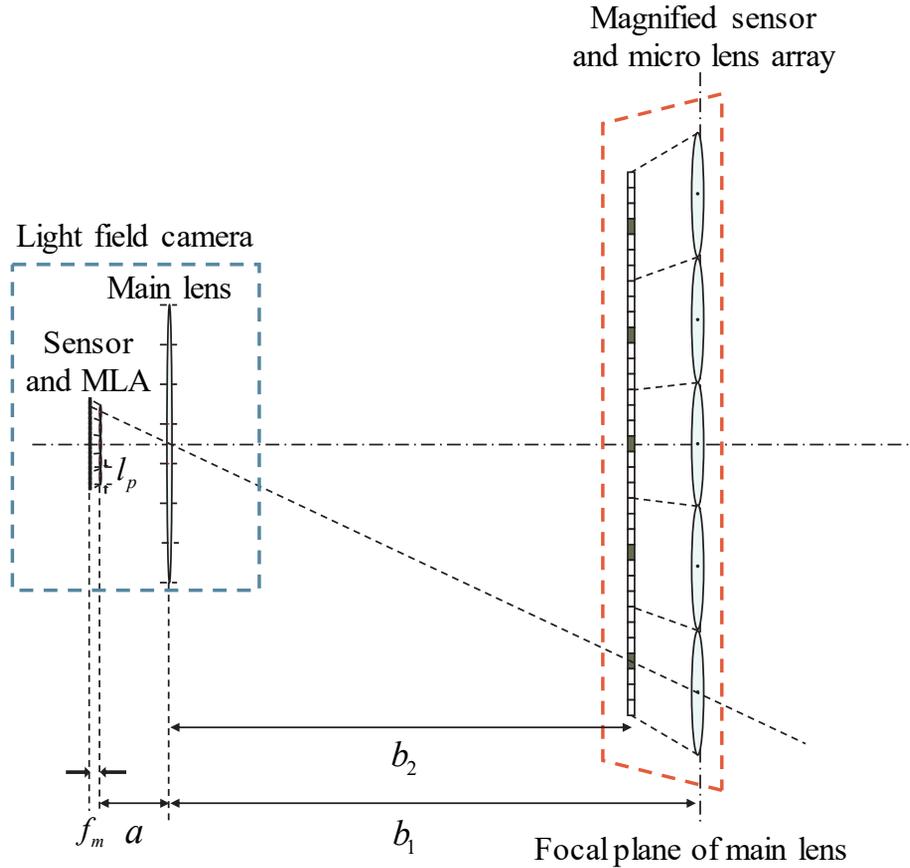


Figure 3.1 Geometric structure of LFC.

3.2.2 Depth extraction method with optical flow for light-field camera

To obtain the depth information of an object by processing the 4D light-field image, depth extraction based on optical flow for LFC is proposed. This method extracts the local disparity between the sub-images using the optical flow method. Optical flow is a distribution of relative motion difference according to the distance between an imaging system and a subject [60]. By finding a corresponding pixel to a given pixel, the method calculates how quickly the pixel flows in the images [61]. Conventional methods of extracting the depth of the object by analyzing the elemental image is the sum of absolute difference (SAD), sum of squared

differences (SSD), and sum of sum of squared differences (SSSD) using the minimum difference function [51, 52]. These methods, however, have problems that the extracted depths are quantized and not able to measure in sub-pixel unit. To solve these problems, a method has been proposed for averaging the depth information extracted by comparing optical flows between sub-images and applying a median filter [57]. However, optical flow method for LFC has not been suggested yet. Captured light-field information of LFC also has very narrow baseline to compare the position of corresponding points [53, 54]. Therefore, sub-pixel-level depth extraction algorithms are essential for the LFC. The corresponding points between the images can be found by comparing the sub-image at the very center of the light-field information and periphery sub-images. According to the specifications of the optical system, the depth information of certain objects is encoded with the disparity between the sub-images calculated by the optical flow value.

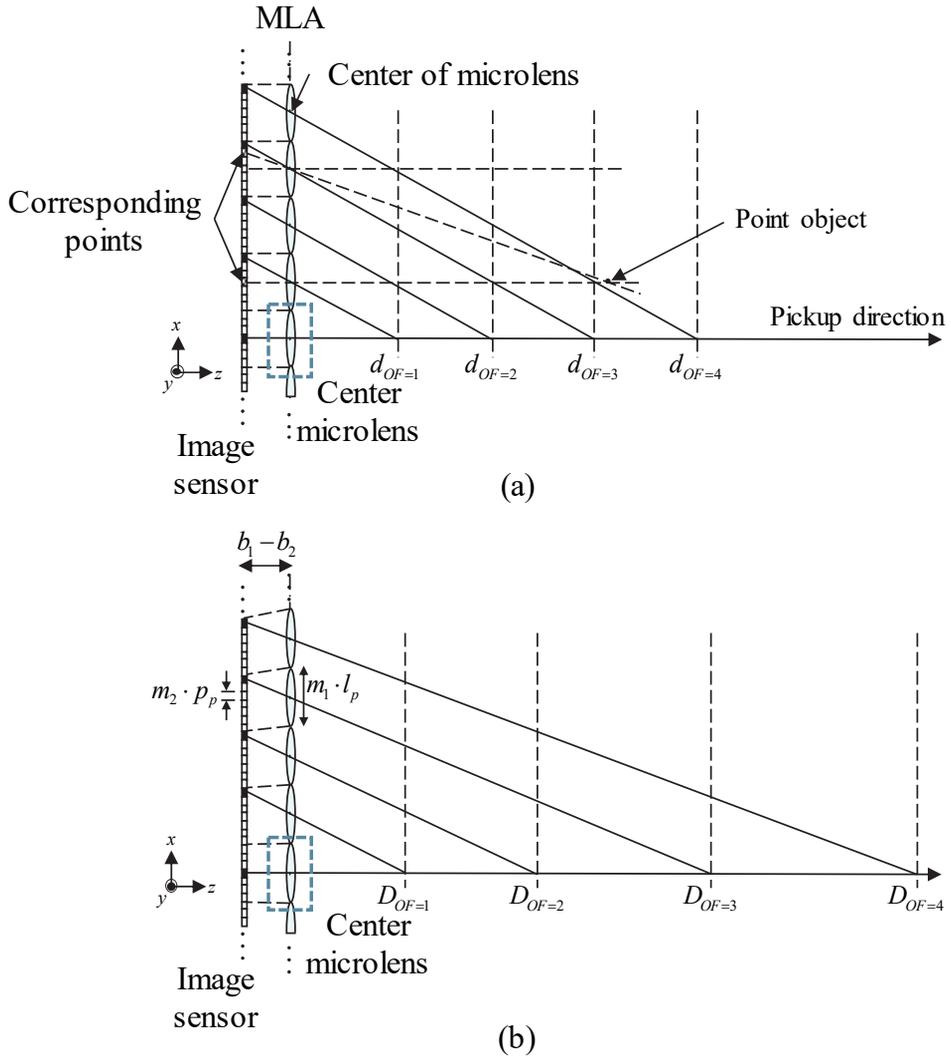


Figure 3.2 Depth extraction from captured light-field using optical flow in (a) integral imaging and (b) light-field camera.

Figure 3.2 shows geometrical relationship between optical flow data and depth information in InIm and LFC pickup system. The OF means the value of optical flow between two corresponding points. The main difference of two optical systems is that, in the focused mode InIm, the axes of each microlens and image sensor are both parallel and perpendicular to the optical system. In case of pickup by the LFC, however, the MLA and the image sensor are differently located and

magnified depending on distances from main lens to the optical elements. Since the magnification values of both optical elements are different, the microlens and the part of image sensors assigned to them are tilted out of the main optical axis depending on their positions. Because two capturing systems have different optical structures, depth extraction method for LFC should be reestablished. It is supposed that lens array in InIm and magnified MLA in the LFC have the same size. The m_1 and m_2 are magnifications of MLA and image sensor, respectively. p_p is the pixel pitch of the image sensor. As shown in checked pixels in Figure 3.2(a), a point object in space can be repeatedly recorded on the image sensor by MLA. The core task of depth extraction is to reveal the relationship between location of repeating points and depth information. In focused mode InIm pickup, for example, let two checked pixels on the Figure 3.2(a) have a corresponding point relation. The rays recorded at two pixels start from the same point but pass different centers of microlenses. Because the distance between two corresponding points on image sensor is two times larger than the lens pitch, the optical flow is calculated as 2. In Figure 3.2(a), the black pixel on center lens image and other black pixels are supposed to be corresponding points to each other. Depending on the optical flow values, the depth of each corresponding point is determined. The position of the depth according to the calculated optical flow is represented as d_{OF} in Figure 3.2(a) and the depth increases linearly with the value of optical flow. In case of LFC shown in Figure 3.2(b), the depth of optical flow value is represented as D_{OF} . Figure 3.2 shows only the integer values of the optical flow for the sake of convenience though this algorithm has sub-pixel accuracy.

The relationship between depth information and optical flow in LFC image should be analyzed to determine DOLF. To solve the depth quantization of extraction algorithm, the accuracy of depth map is secured by applying a median filter or averaging process to each depth map obtained by comparing each sub-image with the center sub-image. Equation (3.1) describes the relationship between

the depth information and optical flow in an LFC. Once a sub-image to be compared and optical flow of corresponding points are determined, the location of the point object is calculated simply by triangle proportionality. Let L be the pitch of the magnified microlens and P be the pixel pitch of the magnified image sensor, i.e. $L = l_p \times m_1$ and $P = p_p \times m_2$. When the resolution of one lens image is $n_x \times n_y$, center index of lens is represented (i_c, j_c) respectively. The distance between two optical elements can be calculated as $|b_2 - b_1|$ by lens formula. In this chapter, it is assumed that optical flow only uses the result compared with the center sub-image. O_h means optical flow in the x -axis, and O_v means that in the y -axis. $O_{h(i,j)}(x,y)$ or $O_{v(i,j)}(x,y)$ denotes a 2D input function where the output is an optical flow value for a given lens index comparing the center sub-image. Note that input variables of O_h and O_v are omitted for simplicity in Equation (3.1). If x and y are coordinates in the center sub-image, then the depth for the center sub-image can be obtained as:

$$D_{OF}(x, y) = \frac{b_2 - b_1}{2} \left[\frac{1}{n_x - 1} \sum_{i=1, i \neq i_c}^{n_x} \left| \frac{O_{h(i, j_c)} \cdot L}{O_{h(i, j_c)} \cdot P \cdot n_x + (i - i_c)P - O_{h(i, j_c)} \cdot L} \right| + \frac{1}{n_y - 1} \sum_{j=1, j \neq j_c}^{n_y} \left| \frac{O_{v(i_c, j)} \cdot L}{O_{v(i_c, j)} \cdot P \cdot n_y + (j - j_c)P - O_{v(i_c, j)} \cdot L} \right| \right] \quad (3.1)$$

The DOLF method is described in Equation (3.1). Using the above equation, one can generalize the case of extraction using InIm technique. In InIm, the lens pitch and the image sensor assigned to it are assumed to have the same size. In other words, in the above equation, $P \times n$ is equal to L , which results in erasing the O -related term of the denominator. The value of $b_2 - b_1$ simply becomes the focal length of the lens array. This result is consistent with previous studies using InIm [57]. The above depth calculation method can be an example of the extraction method using optical flow. In the above equation, the optical flow for the center

sub-image is calculated in the horizontal and vertical directions and the values are averaged. Depth maps obtained from several sub-images are averaged again. However, depending on the situation and user preferences, various statistical techniques other than averaging can be adopted, such as the median filter.

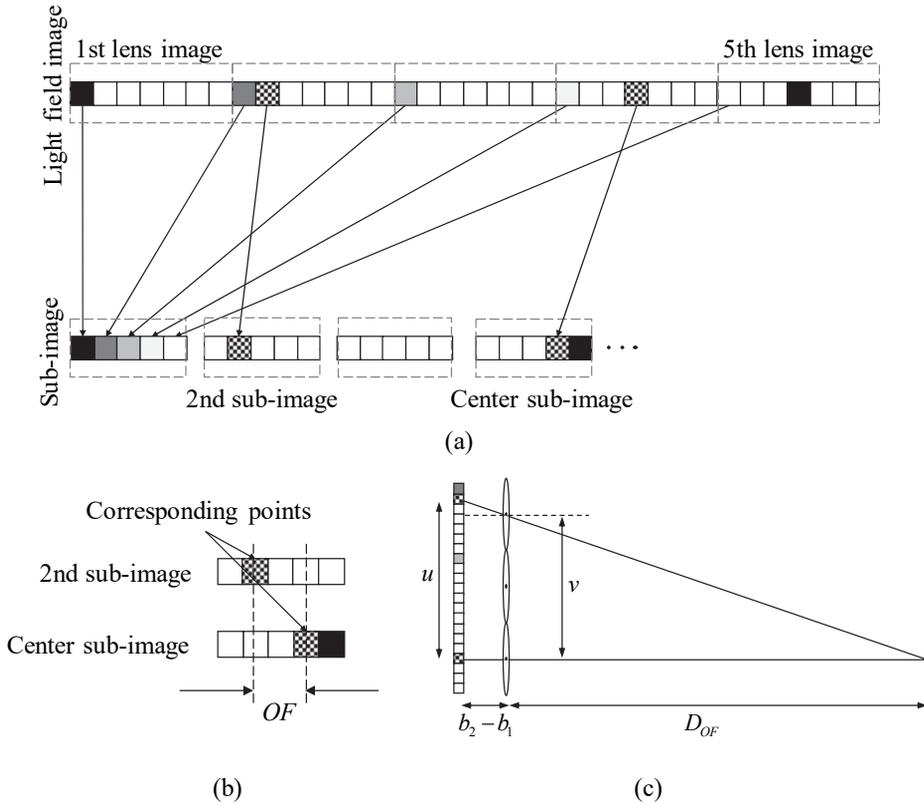


Figure 3.3 (a) Sub-image extraction from light-field image, (b) optical flow calculation between corresponding points, and (c) schematic diagram to derive Equation (3.1).

The Figure 3.3 shows a more detailed description of how to implement DOLF and introduces the derivation of Equation (3.1). First, Figure 3.3(a) conceptually shows the process of extracting the sub-image used for optical flow calculation from the photographed light-field image. The proposed method is most commonly used in a variety of applications using lenslet arrays, for example integral

photography, integral imaging, and light-field microscopy. As can be seen from the light-field image, there are five microlenses, and seven pixels per lens are assigned on the image sensor. It is assumed that the same shaded pixels have the same brightness, and two pixels of the checkered are corresponding points to each other. Sub-image creation is the process of rearranging the pixels of the lens images that are in the same position relative to the index of the lens to which they belong. Looking at the first sub-image on the far left, one can see that the leftmost pixels of each lens image are gathered together to form the first sub-image. The checkered pixels are each placed at the second pixel of the second sub-image and at the fourth pixel of the central sub-image (fourth). After the sub-images have been extracted, Figure 3.3(b) shows the process of finding the corresponding points between the two sub-images and obtaining the distance between them by using the optical flow algorithm. In general, the output of optical flow does not have its own unit, and it is necessary to fit the output value according to the situation to suit the actual situation. In this dissertation, the optical flow values obtained from two sub-images have the pixel width in units, but there still exists area for further development. First, depending on the indices of the sub-images used, which mean their relative positions, the distances of corresponding points become different. Second, the optical flow algorithm can extract the positions of two points with sub-pixel accuracy, although the optical flow value is described as having only integers. For these two reasons, one needs to analyze how the output optical flow value has a certain distance in a given situation.

The final step to derive Equation (3.1) is shown in Figure 3.3(c), which is a more simplified illustration of Figure 3.2 in the text. Given the details of the LFC and the information about the focal plane, one can see the size of the magnified sensor and microlens. The distance between two optical elements can also be obtained, denoted by $b_2 - b_1$. A variable v should be considered to get the size of D_{OF} . The v is the distance between the microlenses where the two corresponding points

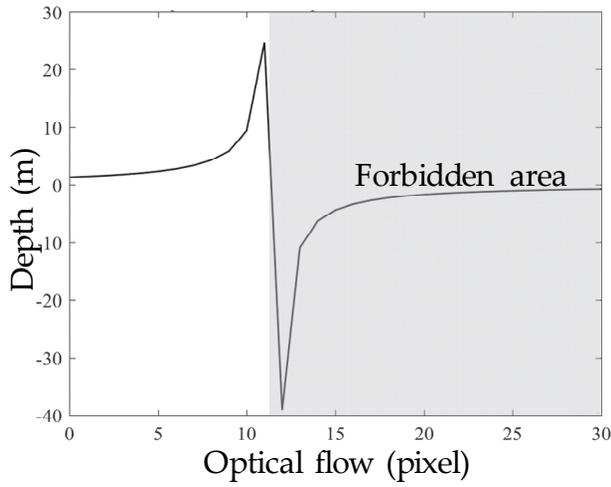
have belonged. From the fact that the position of each pixel in the sub-image is equal to the index of the lens image where it is included before, one can infer that v is related to the value of optical flow. For example, if the optical flow value is 2, the two pixels come from a lens image with an index difference of 2. Therefore, v is a value directly related to the optical flow, specifically, $v = L \cdot OF$ where OF means the value of optical flow and L is the pitch of the microlens considering the magnification. On the other hand, u represents the distance between two pixels on the floated image sensor. The distance between two pixels must be considered – both the distance between the lens images and the positions at which the corresponding pixels lay in the lens images. If the number of pixels allocated to one microlens is n and the size of the magnified pixel is P , then the distance caused by the optical flow difference is $P \cdot n \cdot OF$ and the term $(i_c - i)$ related to the index difference of the sub-image is added, where i represents the index of the sub-image and c represents center. As a result, by simple geometry, D_{OF} can be expressed as:

$$D_{OF} = \frac{(b_2 - b_1) \cdot (O \cdot L)}{O \cdot P \cdot n + (i_c - i) \cdot P - O \cdot L} \quad (3.2)$$

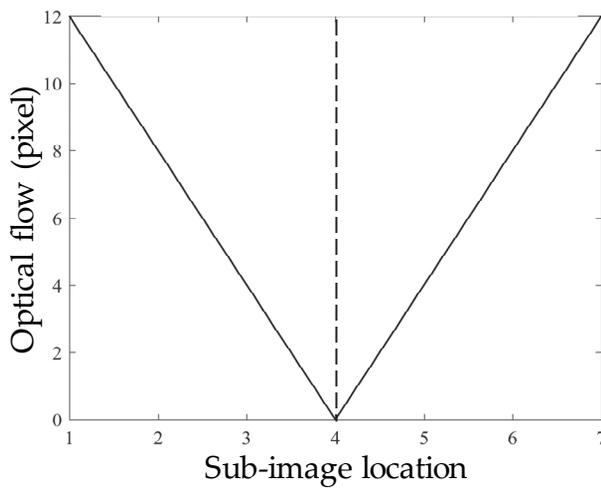
which can be derived to Equation (3.1) by averaging on horizontal and vertical directions according to indices.

There is another difference between two pickup methods shown in Figure 3.2. In 3D object pickup using focused mode InIm, there is no theoretical limit on the maximum value of optical flow for captured light-field. If the sensor format of pickup system is large enough, that is, if the overall size of the lens array and image sensor are large enough, the maximum optical flow value that can be captured also increases. This is because the central optical axes are still parallel, even though a microlens is located at the periphery. Thus, the pixels located near

the edges of lens images can still converge at a physically plausible location. On the contrary, a certain optical axis of microlens and the main axis of LFC converge at a physically impossible position because the optical axes of microlenses are tilted out of the main optical axis according to their positions. As the optical flow value increases, the extracted depth gradually increases and soon diverges as shown in Figure 3.2(b). This phenomenon occurs at a smaller optical flow value when nearer sub-image to center sub-image is used. In other words, the available optical flow range is determined according to the view interval from the center sub-image.



(a)



(b)

Figure 3.4 (a) Extracted depth of center sub-image with seventh sub-image (among seven sub-images) according to optical flow (b) Maximum optical flow on location of sub-image (among seven sub-images).

Figure 3.4(a) represents the depth on optical flow value by DOLF for corresponding points on center sub-image and another sub-image. Since two pixels that converge above a certain optical flow are located at the position that cannot be physically allowed, it is necessary to use a mask to neglect a range of optical flow

value in forbidden area. In this chapter, the mask for the specified range of optical flow value is applied to a case where the convergent depth position is completely unrealistic, that is, when the depth information is negative. Figure 3.4(b) shows the maximum optical flow value that is available for the center sub-image in the order of the sub-images. The dotted line indicates the middle index of sub-image. The optical flow values shown here are absolute values. Though, for example, the third and the fifth sub-image have the same optical flow value, the calculated data flows in different directions. Figure 3.5(a) and (b) show simulated light-field image and Figure 3.5(c) represents the result of depth information extracted from light-field by optical flow calculation using Equation (3.1). The raw light-field photographed by the LFC and extracted the light-field information through the modified optical flow method is simulated. The specification for light-field image is presented in Table 3.1. Figure 3.5(d) represents detailed depth information along the white dotted line in Figure 3.5(c). Figure 3.5(d) compares the results obtained by DOLF and optical flow for InIm with ground truth data. To extract a precise depth, it is necessary to consider the geometrical arrangement of the LFC as shown in Figure 3.5(d). This comparison is meaningful when considering that depth extraction algorithms for LFCs extract relative depth rather than actual depth. When extracting the relative depth, it is necessary to fit the extracted value to the actual distance according to the many parameters of imaging system. In this case, the optical configuration of the LFC must be fully considered to obtain the correct depth extraction result. As a result, the proposed depth extraction method for LFC is verified. It is noted that the data of ground truth is more discontinuous than the result of depth extraction on the center of graph because the original depth map of sub-image has been quantized after 8-bit conversion process.

Table 3.1 Specification for simulation and experiments.

Elements	Specifications	Value
Microlens array	Lens pitch	38.4 μm
	Focal length	77 μm
	Number of microlenses	501 \times 501
Main lens	F-number	2
	Focal length	50 mm
Imaging system	Image plane distance (a_1)	52.08 mm
	Object distance (b_1)	1.25 m
	Sensor pitch	5.5 μm
	Total sensor resolution	3507 \times 3507
	Lens image resolution	7 \times 7

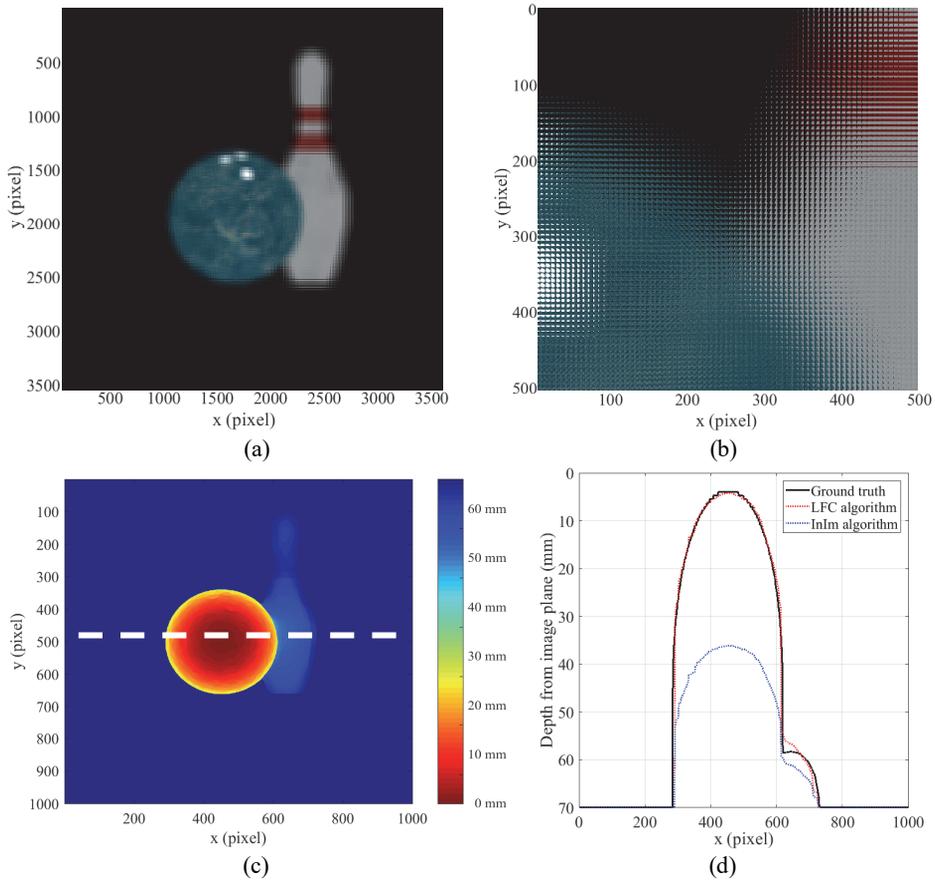


Figure 3.5 Depth extraction example: (a) Example of light-field image, (b) magnified light-field image, (c) depth extraction results, and (d) comparison of the ground truth and extracted depth maps obtained by InIm method and LFC method.

3.3 Simulation and analysis for one-shot 360-degree light-field capturing

3.3.1 Analysis for the 360-degree light-field capturing

4D light-field information expressed as a 2D array form is simulated according to the LFC condition. From the light-field image, the depth extraction adopting optical flow method is introduced. Based on these studies, a system for acquiring 360-degree 3D information of object using LFC at a single capture is proposed. The simplest structure to achieve our purpose is chosen. The proposed system consists of one LFC and a couple of mirrors. The light-field information obtained by the system depends on the arrangement of viewing angle, mirrors, and objects. With two mirrors and one LFC, three perspectives surround and capture the object according to their locations and orientations. Since the three light-field images need to be combined to restore 360-degree information of the object, an analysis of the location and orientation of the optical elements should be preceded. For analytical simplicity, the structure of the proposed system includes the two following assumptions. First, the two mirrors are large enough to reflect a given object. Second, three perspectives have equal usage of viewing angle on image sensor. This is to ensure that the light-field information taken at three viewpoints equally includes the angular information of the object.

Figure 3.6(a) shows the overall proposed pickup system. The LFC is oriented to the front of object and two mirrors are aligned to compose two additional perspectives as described above. Therefore, the image sensor in LFC records three perspectives simultaneously. It is supposed that three light-field images laid on sensor have same area on sensor. In other words, a third of image sensor is allocated to each perspective horizontally. One of the perspectives uses the field of view (FoV) θ , when the total FoV of LFC is 3θ as shown in Figure 3.6(b). Figure 3.6(b) only shows left part of the system for less complexity. The point S is the center of object and the point C means the center of main lens of LFC where the

center of FoV exists. The point S' is an image of the point S reflected by the mirror and the distance between the mirror and the point S is l . The distance from the point S' to main lens is represented as s . To make sure three perspectives have the same angle differences, two green dotted lines should form an angle of 120° as shown as Ω in Figure 3.6(b). The location and orientation of the mirror are determined by two factors, the length of perpendicular line from point S to point C , represented as d and the angle between optical axis of LFC and mirror surface, φ . From geometric analysis, the angle φ can be simply calculated as $\Omega/2-\theta/2$. The length l and s are determined by following Equations (3.3) and (3.4).

$$l = \frac{\sin \theta}{2 \cos \left(\frac{\Omega + \theta}{2} \right)} d \quad (3.3)$$

$$s = 2l \sin \varphi + d \quad (3.4)$$

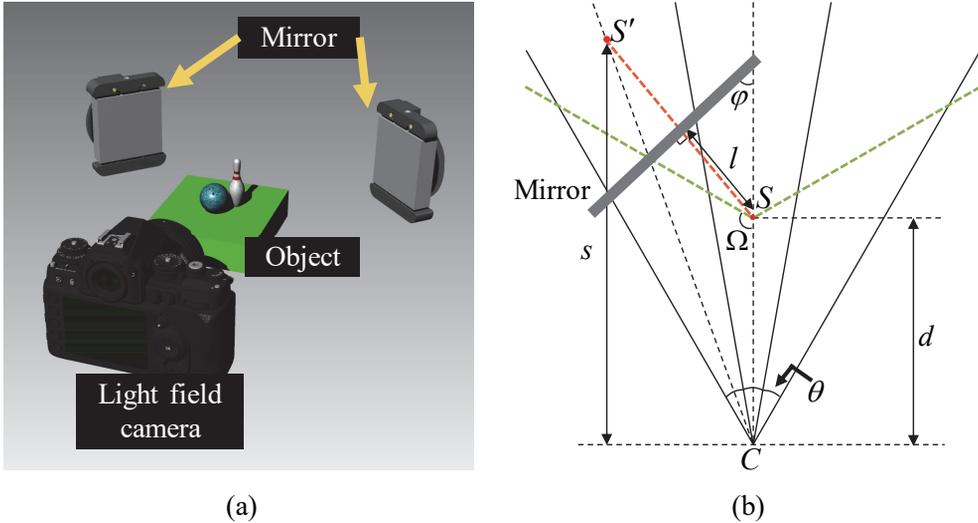


Figure 3.6 (a) Proposed light-field pickup system and (b) geometrical figure for analysis on the proposed system.

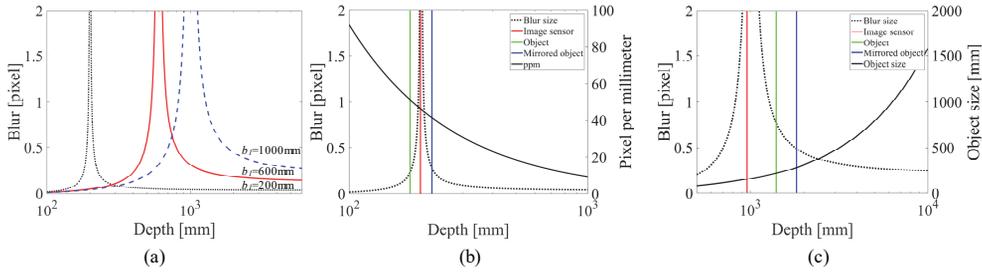


Figure 3.7 Blur size depending on depth of images. (a) Blur radius of microlens versus depth of scene according to various focus distances (b_1) in log scale. (b) Focal plane between the object and the mirrored image: focus distance (b_1) is 200 mm and object location is 180 mm. Solid black line represents spatial resolution of captured image in pixel per millimeter. (c) Focal plane in front of object: focus distance (b_1) is 1000 mm and object location is 1400 mm. Solid black line represents maximum object size to be captured.

To simultaneously capture images in S and S', it is necessary to analyze where the focal plane of the LFC is located. When a regular camera focuses on a specific position, the depth of field (DOF) is determined based on the specification of the camera and the information of the object can be acquired without blur in that range. On the other hand, in the case of the LFC, the size of blur according to the distance behaves in the opposite manner to blurring mechanism of regular camera. The blur made by LFC appears small in most areas except the near the focal plane, and the size of blur is bounded to that of the microlens in the focal plane. Analysis of blur for LFC has been done in several researches [4, 28, 30]. Based on these previous research, the blur radius of microlens is calculated according to the distance to various focal planes which is shown in Figure 3.7(a). The focal lengths of the main lens and the MLA are 50 mm and 0.12 mm, respectively. When the focal plane moves away from the camera, the image of the microlens becomes thick in space, and therefore the light-field information of the object placed near the focal plane may not be captured properly. This suggests that the method of finding the optimal focal point depends on the position of the object. According to Equations (3.3) and

(3.4), the position of the mirror image can be shown in Figure 3.7(b) and (c) depending on the position of the object. When the object is relatively close to camera, for example 200 mm in Figure 3.7(b), the space suffered by blur is small enough to be between points S and S'. In this case, the depth of the object can be obtained properly, except when the peak of the blur covers one of the two images. However, the farther away from the peak of the blur, the greater the number of repetitions of a point on the light-field image, which leads to the possibility that both perspectives will invade their respective regions. Therefore, in the case of Figure 3.7(b), it is desirable that the peak be located between two objects. When the object distance is relatively large as the case of Figure 3.7(c), it is impossible to place a focal plane between two objects because the width of the peak is too thick. In this case, the peak should be placed on either side of the two objects. However, as the distance from the camera increases, the number of pixels used for recording per unit distance decreases which is shown as pixel per millimeter (ppm) in Figure 3.7(b). Therefore, to record two objects in similar quality, it is desirable that the near object undergoes a larger blur. Consequently, in the case of Figure 3.7(c), the focal point of the camera should be in front of the object. Meanwhile, the comparison between black solid lines in Figures 3.7(b) and (c) shows the tradeoff between the size of the photographable object and the amount of information per unit length of the recorded light-field. Therefore, it should be placed at a distance from the camera that achieves the highest resolution depending on the size of the given object.

Meanwhile, mirrors having a specific curvature can be used in the proposed capturing system. Since the mirrors in our system do not magnify images, the reflected images are located beyond mirror and occupy smaller portion of the image sensor, compared to the center perspective light-field image. To enhance light-field resolution on the sensor, a pair of optical elements which magnify an object can be used, such as concave mirrors. When one uses concave mirrors to the

proposed system, the location and orientation of the mirrors are determined by the distance between main lens and the object, like plane mirrors. Also, the focal length of the concave mirror adjusts the location and size of magnified image. Even though maximum magnification is controlled by the viewing angle of the LFC, there is a possibility that the overall shape of object limits magnification value. There is another practical consideration of using concave mirrors. Because longitudinal magnification in imaging system is not linear in contrast to transverse direction, depth information of magnified images should be manipulated after DOLF algorithm. A shape of object should also be considered to optimize the overall system. If the object has many wrinkles and occluded parts, some parts of object cannot be captured by an LFC. In this case, multiple mirrors which are located at several positions toward the object are necessary. There may happen another issue to use the proposed system. Though 3D scanning is the main application of the system, independent objects also can be captured on the proposed system at the same time. However, if there are more than two objects or an object which has severely folded structure, some part of the 3D object cannot be recorded and retain occluded parts. There are several researches about reconstruction of occluded object in the field of InIm but the occluded part to be restored is quite limited according to the size of the object [57]. Using additional mirrors can be an effective solution to record entire space around objects. If the image sensor is large enough to accommodate resolution for assigning more mirrors, mirrors on the several locations can solve this limitation. If an optical configuration that makes three perspectives overlapped, occlusion removal algorithm also can be used in this case. After sectioning three objects according to the depth of the light-field image, texture information of the occluded position is reconstructed by using a combination of sub-images.

3.3.2 Four-dimensional light-field simulation using sub-image synthesis

To carry out an optical simulation for an LFC, it is necessary to define the optical elements that characterize the LFC. The 4D light-field information to be captured is distributed differently according to the focal length of the main lens, the focal length and apertures of the MLA, the pixel size of the image sensor, and the overall resolution on light-field plane. When the LFC is focused at a certain depth to capture an object, the distance between the MLA and the main lens changes. According to this distance, a different magnified image sensor and MLA in space capture objects. The MLA, main lens, image sensor and the depth of focal plane determine the specifications of the magnified pickup unit. One of methods to make 2D form of light-field is to simulate a sub-image which can be extracted from the target light-field. A sub-image can be created by collecting pixels in the same position in each lens image. Before taking a sub-image, the orientations of pixels in rendering software should be determined according to the specifications of LFC [50]. From given details of LFC, virtual cameras for sub-images are made in space as the form of Figure 3.1. The resolution of the camera is same as the number of microlens and the distance between pixels is lens pitch of magnified MLA. From the simulated sub-images, one can restore the light-field by relocating pixels of sub-images. The center image of Figure 3.8(a) shows one of the light-field images by an LFC. Total 49 sub-images, which comprise 7×7 images in horizontal and vertical direction, are merged to make light-field. Resolution of each sub-image is 500×500 . The specifications of the camera used for the simulation are given in Table 3.1.

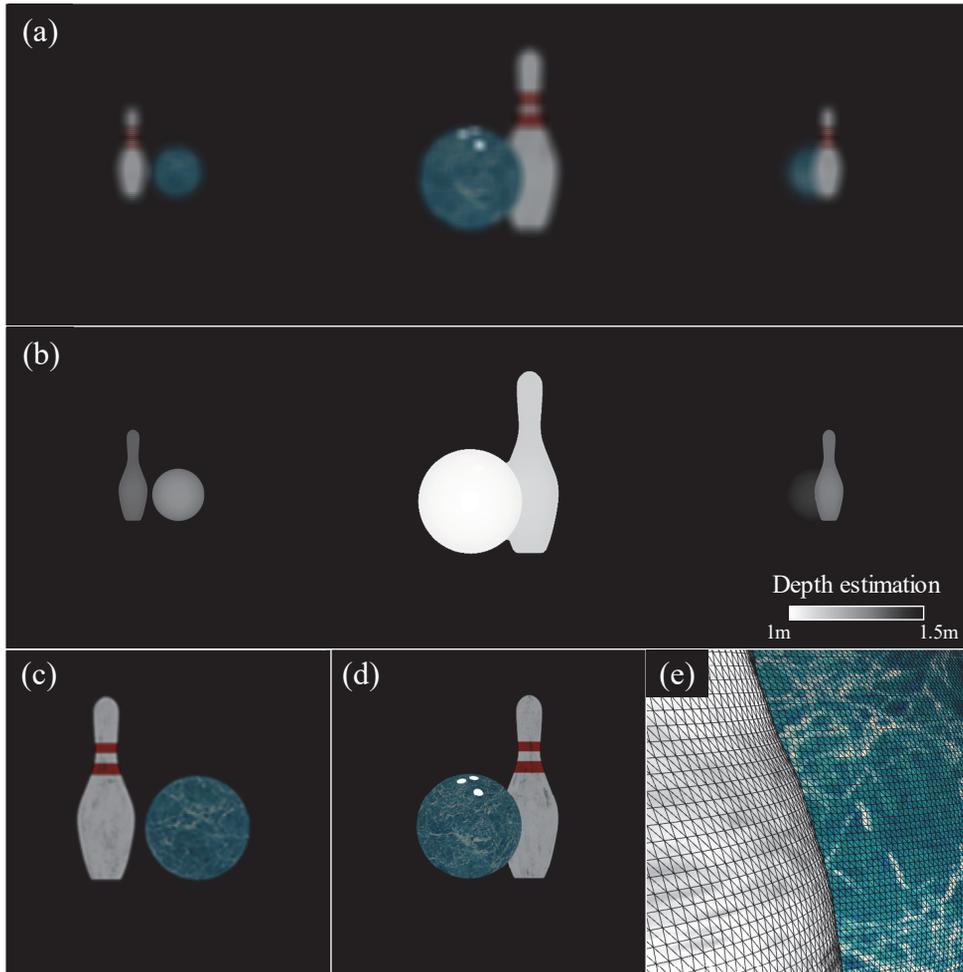


Figure 3.8 Simulation results for 360-degree light-field capturing and DOLF. (a) Simulated light-field image in the proposed optical setup, (b) extracted depth information of the light-field, (c) refocused image of left perspective of the light-field, (d) refocused image of the center perspective, and (e) reconstructed 3D triangular mesh from the color point cloud.

To verify the optical system configuration based on the above analysis, optical simulations are performed with virtual mirrors, LFCs, and objects. POV-Ray software is used to place the 3D information of the object and to make the light-field, and depth extraction method was performed using Matlab software. By using the LFC simulation, light-field in the proposed optical capturing system is

generated. Figure 3.8(a) shows the resultant light-field. The specification of MLA and main lens are also presented in Table 1. The overall condition of pickup system is also the same as the table. However, the focal plane of the main lens is slightly moved back to the camera to represent light-field effectively. As shown in the Figure 3.8(a), two additional viewpoint images are taken using a flat mirror. Because the mirror does not have any magnification, the reflected image has a smaller resolution than the front image. As mentioned above, concave mirrors can be an alternative to improve resolution of reflected light-field images. With the optical flow algorithm for LFC, the depth information of the simulated light-field is also extracted as shown in Figure 3.8(b). Note that the depth information is normalized to 8-bit information within proper depth range according to the size of the object. To record the texture information of the light-field, digital refocusing technique for LFC is used [4]. As shown in Figures 3.8(c) and (d), color information of each perspective is demonstrated. Since the resolution of reflected light-field has lower resolution than that of center perspective, Figure 3.8(c) looks blurry compared to the Figure 3.8(d). The three pairs of resultant depth map and color information can be projected in virtual space as a form of point cloud having 3D position and color information. To combine the three light-field images, the mirrored perspectives should be resized after digital refocusing. The magnification ratio between center and the others are determined by $s:d$ as shown in the Figure 3.6(b). According to the ratio, the right/left sub-images are resized to have same scale in the results. By binding all point cloud data, the 3D object can be reconstructed finally. To transform the point cloud data to solid one, a triangular mesh is one of the alternatives to handle 3D information. Various algorithms to make mesh type object from point cloud data have already been suggested [57, 62, 63]. Figure 3.8(e) represents reconstructed triangular meshes of the object, which are rendered from the back of the 3D object.

3.4 Experimental results for the light-field capturing and reconstruction

3.4.1 Experiments for one-shot 360-degree light-field capturing

An optical structure for 360-degree light-field capturing was implemented as shown in the Figure 3.9(a). Two plane mirrors and one LFC were used to compose the overall system. To align each optical component, positions of LFC and stage of the subject is adjusted firstly. The positions of two mirrors were controlled with a linear translation stage. A simple laser pointer is utilized to adjust orientation of the mirrors. With the composition of the optical elements, the position for subjects was fixed on the stage. The stage for the subject and background of the studio were wrapped with solid color sheet, which was green color in our system. To remove the background data, chroma-key technique was used. Figures 3.9(b) and (c) show the captured light-field data and their depth map information, respectively. The background noise was removed in the raw light-field and the light-field was cropped to calculate accurate depth map. The resolution of processed light-field was 3864×2681 . Since the resolution of each lens image was 7×7 , the resolutions of each sub-image and depth map were 552×383 .

Figures 3.9(d) and (e) show left and center perspective images, which were digitally refocused to extract clear color information. In the comparison of Figures 3.9(d) and 8(e), the center perspective had the higher resolution than left and right one. The original resolutions of the center and left or right perspective were 1020×1020 and 510×510 , respectively. To acquire proper refocused images, involving enough DOF was essential according to the size of subjects. Since the f-number of the light-field was 2, which provided rather low DOF for most of objects, a synthetic aperture technique whose f-number is to reduce converging rays onto sensor and increase DOF of the synthetic image is used. Figure 3.9(f) shows 3D color information of red box region in Figure 3.9(e). Triangular mesh

was used to illustrate the 3D structure and the location of virtual camera for rendering was adjusted to emphasize the peak of the penguin. As to the simulation results, the color point cloud was located on virtual space in the PC environment, bound altogether, and reconstructed as 3D meshes.

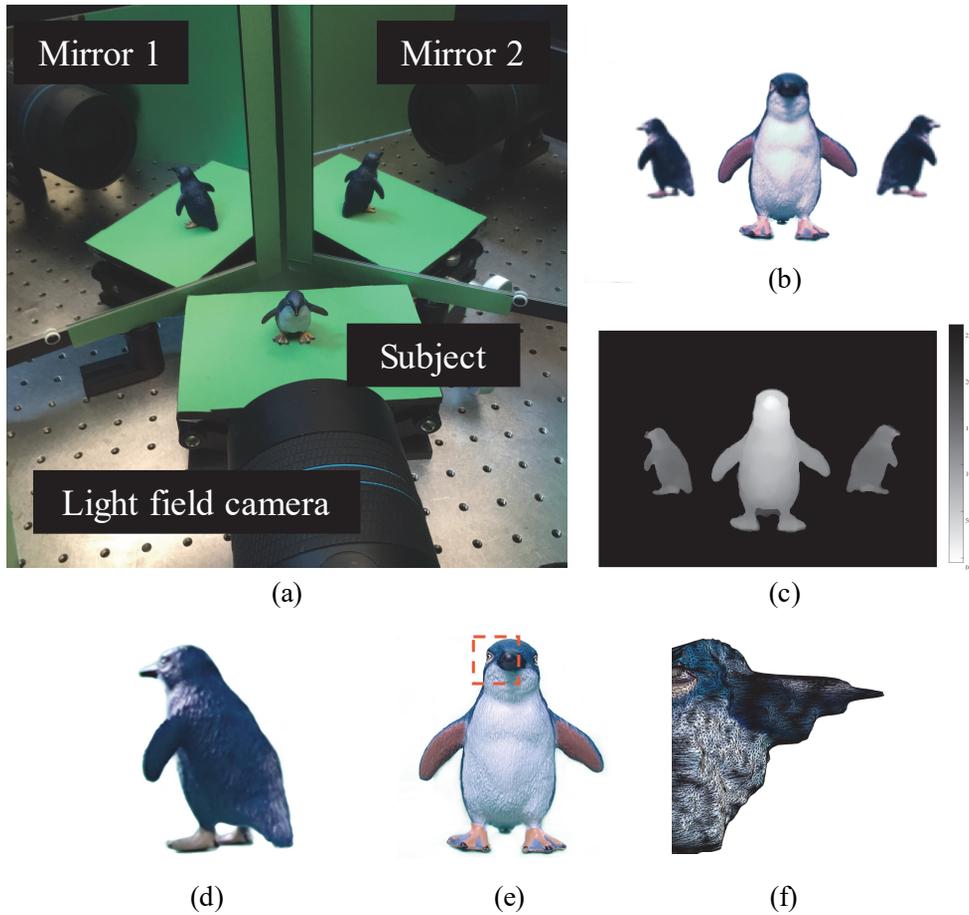


Figure 3.9 The 360-degree light-field capturing setup and experimental results. (a) The implemented capturing system with two plane mirrors. (b) Captured light-field image focused on the center perspective and processed with chroma-key technique. (c) The depth map of the light-field calculated by the proposed algorithm. (d) Refocused image for left perspective. (e) Refocused image for center perspective. (f) Reconstructed mesh with color information.

3.4.2 Experiments for optical reconstruction of the light-field

Not only to capture 3D information with the proposed method but also to display 3D object optically, holographic display to reconstruct 3D information of the proposed capturing system is adopted. Figure 3.10 shows top view of holographic display system for verifying reconstructed 3D information. Fast spatial light modulator (SLM) was used to display holographic information. The SLM, which was implemented by a digital micro-mirror device (DMD), could control the amplitude of the incident wave by toggling micro-size mirrors very quickly. The DMD was V-9501 model of Vialux. In the principle of Fourier hologram, an amplitude hologram can be reconstructed to a complex hologram. The computer generated hologram can be made from intensity and depth information of each pixel. To enhance quality of the reconstructed hologram, random phases were wrapped on each frame of the hologram and timely multiplexed [64]. Because the aspect ratio of DMD was not square, reconstructed hologram in the Fourier plane shrank along horizontal direction without compensation. Anamorphic Fourier optical system, which was composed of three cylindrical lenses in Figure 3.10, was implemented to optically correct the distorted pixel structure and image [65]. The focal lengths of cylindrical lens 1 and 3 were 200mm and the focal length of cylindrical lens 2 was 100mm. The distance between adjacent cylindrical lenses was the focal length of cylindrical lens 2. Figure 3.11 shows accommodation cue of the reconstructed holograms. The letter of front and rear were located on front and rear positions of the objects.

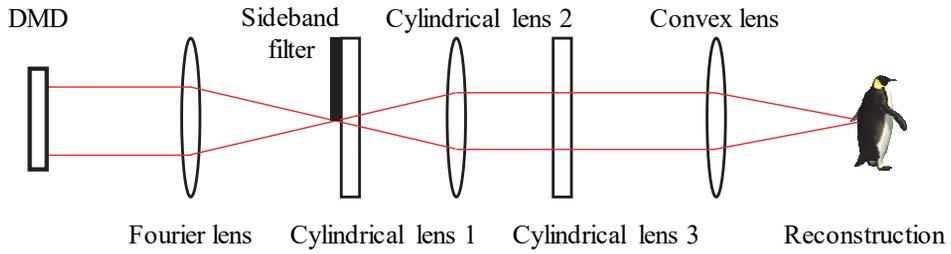


Figure 3.10 Top view of the holographic display system with anamorphic optical transformation model.

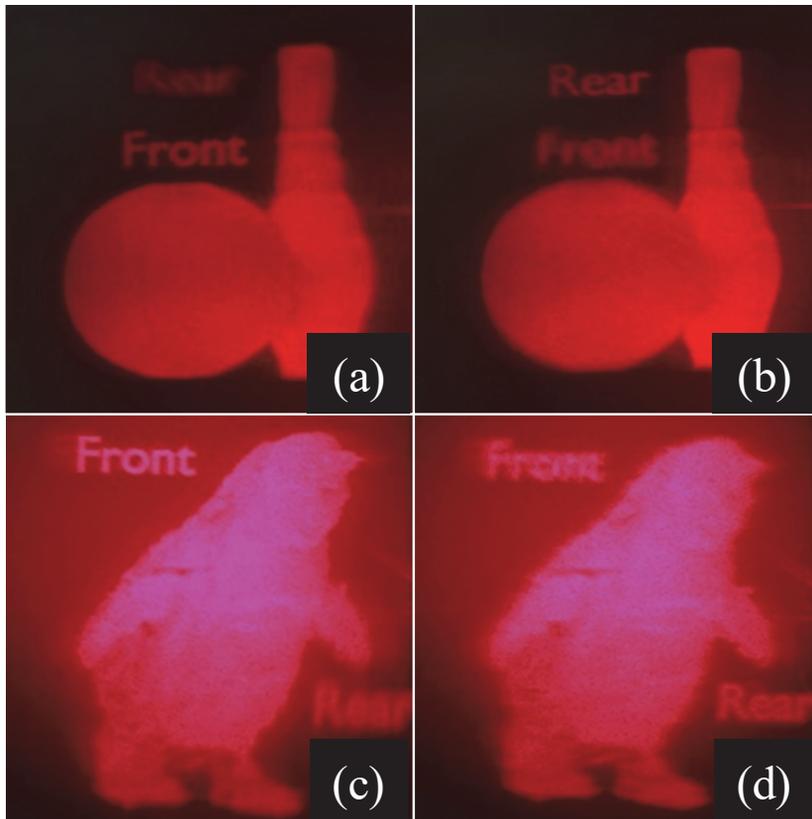


Figure 3.11 Optical reconstruction of captured information with holograms. Captured hologram of simulated light-field focused on (a) Front and (b) Rear. Captured hologram of the captured light-field focused on (c) Front and (d) Rear.

3.5 Conclusion

In this chapter, a system for capturing 360-degree light-field information of one object in one shot and optically reconstructing it is proposed. To process 2D light-field image to depth profile with texture, depth extraction method based on optical flow for light-field camera (DOLF) is also introduced. Maximum optical flow value and a mask for the value are introduced according to relative position difference between sub-images in light-field. For 360-degree capturing, a compact and inexpensive studio composed of two mirrors and an LFC are used. As a result, light-field information around a subject is captured in one-shot. Experimental and simulation results are presented to support the proposed system and analysis. The resolution of the used light-field is 3864×2681 and the resolution of each sub-image and depth map is 552×383 . Holographic display with anamorphic Fourier optical system and DMD is used to optically reconstruct the captured light-field information.

Chapter 4. Ray-based end-to-end pipeline for optimization of optics and image post-processing

4.1 Introduction

Recording of the information of the external environment in the form of images has been accepted as a couple of separate procedures even after the digital age. Therefore, the main elements of each stage have been designed and refined to achieve the best results within each procedure. First, the light incident through the lens is modulated by the refractive index difference between the surface of the lens and the air, and is transmitted to the surface that can measure the energy of the incident light for a given time period, which is made of a photosensitive film or a pixelated array of photodetectors. In the case of a digital image sensor, various types of image processing steps are required to convert the energy of the acquired light into an electric signal, which is then converted to a form that can be accepted by humans. Starting with demosaicking caused by the use of a general Bayer filter, procedures such as gamma correction and color space conversion have been established over time to improve the non-linearity of the sensor. For final output images in compressed form such as JPEG, image post-processing techniques are also introduced to improve image quality, such as noise reduction, that has not been done in the image processing pipeline. Also, image modulation, such as image deblurring or depth extraction algorithms, has been developed to overcome limitations of conventional optical devices. Recently, research on image processing using state-of-the-art technologies including deep learning has been widely carried out, and introduced to improve the image quality for various type of input images.

Independent and sequential procedures for creating images are often complementary. In particular, if the modulation characteristics of the optical

system are not uniform, the image processing pipeline may include additional algorithms to compensate for this, such as vignetting correction or lens distortion correction. However, this merely compensates for defects in the optical system through image processing and cannot help to design the optical system itself. On the other hand, image post-processing algorithms that perform specific roles by engineering the impulse response of the optical system, that is, point spread function (PSF) information, have been studied using a lens or a diffractive optical element [6]. Optics are optimized for PSF engineering, and optimized optics have produced meaningful results for extended depth of field (DOF), defocus and deblurring, and full-color imaging with DOEs [6, 66]. However, such a PSF engineering method independently dealt with the process of selecting a PSF and restoring the photographed image. In other words, it is difficult to explain the effect of the PSF itself on a specific image restoration algorithm sufficiently clear.

Recently, a study has been proposed to jointly optimize the optical parameters and the image processing algorithm parameters of a lens surface, an image sensor, and an algorithm for restoring an image using a deep learning library as a differentiable pipeline [5]. Since the end-to-end optimization framework simultaneously optimize the reconstruction algorithm and the surface of the lens, it overcomes the disadvantages of the existing independent image processing pipeline and constructs a domain-specific lens and image restoration algorithm for a given computational camera. The pipeline uses a wave-based image formation model using Fresnel propagation to implement this framework using a PSF of the lens in a fully-differential manner. The wave-based image formation method helps to make the overall pipeline differentiable because it can easily obtain the complex amplitude of closed-form when a given input light reaches the image sensor through a specific lens. However, diffraction-based image propagation has a fundamental disadvantage that the maximum diffraction angle is limited by the pixel pitch of the diffraction element. Regardless of surface type of the optimized

lens, for example, an etched glass, or a diffractive optical element, the lens results in an aliasing result for light that must refract over a maximum diffraction angle for a given incident light. This limitation severely restricts the numerical aperture values that an optimized lens can have, and consequently represents the difficulty of present commercial-grade lens designs through wave-based image formation models. Therefore, the recent lens design uses a ray-based simulator, and also receive help from several Monte Carlo methods to solve the sampling problem [67].

In differentiable pipeline, an SGD algorithm is employed to update trainable variables. The algorithm calculates gradients of final loss according to the infinitesimal increments of variables. Since the gradients should be tracked by the SGD algorithm, the overall pipeline is constructed with differentiable image processing procedures. The algorithm changes the trainable variables to reduce the loss of the pipeline with certain amount of step, which is called learning rate. The learning rate can be designated by the users according to the general order of the variables. Since the numerical values of variables might be considerably different, different learning rates can be used to optimize each variable. Several SGD algorithms are represented along the tremendous explosion of deep learning research. Among them, Adam optimizer (Adaptive momentum estimation) is employed to optimize the pipeline, which is broadly applied to image processing now.

In this chapter, ray-based end-to-end optimization for optical element and image processing is proposed. The surface profile of optical elements are encoded with several parameters by using analytic polynomial basis. By using a ray tracing method, each step of ray propagation is implemented as a fully differentiable operation. The optimization pipeline contains a ray tracing part to calculate a PSF of certain optical element, an image pipeline for giving noise on acquired images and resampling information to suppose a certain specifications of image sensor and image post-processing to reconstruct acquired images according to the design

objectives of the optimization pipeline. Parameters for optical elements and image processing is updated while optimized with SGD. To design achromatic optical element, representative refractive indices are considered to reflect wavelength-dependent effects. Also, the ray-based optimization pipeline supports much larger numerical aperture for designed optical elements. For optimizing a lens with high numerical aperture value, the several PSFs for the lens surface are calculated according to incidence angles to take spatial-variant property into account. By using the ray-based optimization pipeline, a computational imaging systems is designed for extended DOF applications. Optimized optical elements are fabricated using diamond turning to verify the simulation results.

4.2 Non-paraxial ray tracing simulation for computational imaging for optimization pipeline

4.2.1 Computation of point spread functions with ray tracing method

To optimize a lens that can perform a specific computational imaging, one needs to know how the lens during the optimization process modulates the input image. This can be tracked by calculating what shape of blur is to be made for a representative input that can reflect the characteristics of the lens - for example, white collimated light or a white point light source at a specific position. The intensity information of the blur is expressed a PSF in this chapter and can be understood as the impulse response of the focused imaging system. The convolution operation is used in the optimization process because the blurred image after passing through the lens can be calculated at one time using the batch inputs. To construct a ray-based optimization pipeline, a ray tracing method is introduced to calculate the PSF of a lens during the optimization process. In order to calculate the PSF of the lens, the surface profile of the lens should be defined firstly. Depending on how the lens is fabricated, the number of pixels that can

express the height of the lens surface is determined. Memory capacity of the workstation can limit the resolution of the height of the lens also. The surface profile of the lens can be optimized pixel by pixel. When pixelated height information is processed, each pixel on the lens surface represents a thickness of the lens material and is subjected to the optimization. For example, optics optimized with 1500×1500 resolution for the fabrication of diffractive optical elements (DOE) will result in 1500×1500 trainable variables.

Treating all height information of the lens as a variable is not only memory consuming but also causes a problem that the optimized result is not a continuous surface which is hard to be produced. Also, depending on the propagation type of pipelines, the pipeline cannot simulate exact PSFs for arbitrary surfaces which can be encountered during the optimization. In conventional ray-based optical element design software, several types of polynomials are used, which is defined by orders and coefficients to define freeform surfaces with arbitrary curves, eg Chebyshev polynomials, and Zernike polynomials. Once the order of the polynomial related to the maximum spatial frequency of the lens surface is determined, the total number and type of coefficients that encode the surface are decided. Compared to DOE optimization, encoding lens surfaces by coefficients is more efficient in that it can reduce the number of variables to be optimized. In this chapter, U use the basis of Zernike polynomial for encoding the lens profile and optimize the surface by updating the coefficients to be multiplied on each basis. Each coefficient of the Zernike polynomial indirectly represents the optical aberration of the wavefront modulated by the parallel light passing through the surface. Therefore, using Zernike coefficients makes it easy to check the properties of the optimized lens while optimizing the surface. For example, the lens to be manufactured in this chapter is basically a focusing optic to transmit external information to the image sensor surface. This focusing property, such as a curvature and focal length, can be tracked by identifying the tendency of the fifth coefficient among the Zernike

coefficients, namely defocus coefficient.

In order to process light information with the ray-based method, input light source is represented by a 4D light-field matrix for its spatial and angular information. Also, wavelength-dependent effects are considered by defining a representative wavelength for each color channel. After calculating the surface normal vector of the refracting surface, angular vectors of refracted light is calculated by using a vector form of Snell's law. In this dissertation, it is assumed that an optical element whose one side is only modulated and the other side is flat like a plano-convex lens. After a first refraction at the optimized surface, the light propagates until meeting the flat back of the lens. The light refracts again toward the air and is transmitted through free space to the image sensor surface. To integrate the light from the image sensor's surface, a two-dimensional histogram is calculated by assuming a pixelated sensor array and counting the number of light arriving at each bin. After converted to appropriate data type and normalized, the histogram can be used as a PSF for convolution and deconvolution computation. According to the position, depth, and angle of incidence of the input light source, the ray-based PSF calculation method can be performed by changing input light sources. Figure 4.1 shows schematic diagram of the imaging system for the proposed computational imaging. According to the target depths and number of ROIs, the point light sources are declared with their position and head for the target surface. One of the surface of the target lens has target surface which is updated while optimization process. The target lens have parameters containing its size, refractive indices for color channel and fabrication resolution. The image sensor is divided into several ROIs and has parameters for total size, pixel pitch and resolutions. The distance between lens and the image sensor is also defined.

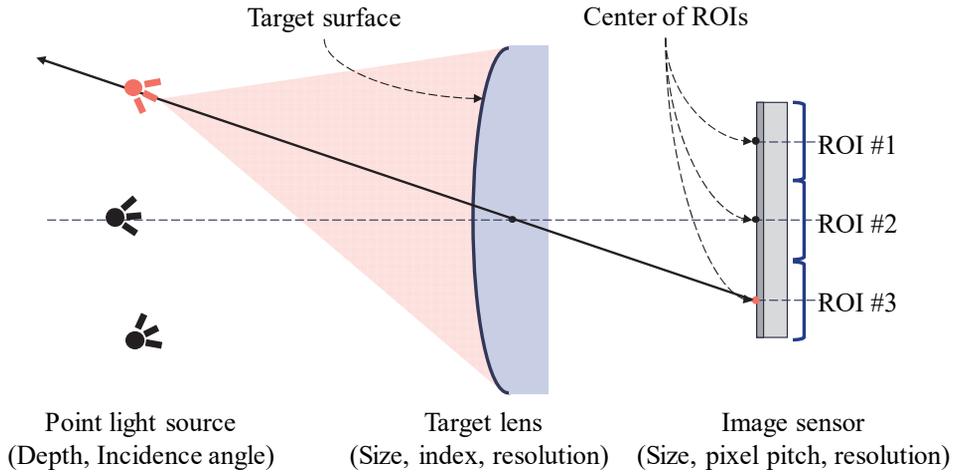


Figure 4.1 Schematic diagram of the imaging system for the proposed computational imaging.

Compared with Fresnel propagation method, ray-based optimization pipeline can support larger numerical aperture of lens by not limiting the angle of refraction. However, a lens with higher value of numerical aperture should be designed in consideration of not only on-axis but also off-axis input light which makes severe lens aberrations according to incidence angles. It means spatial-invariant property of PSF is not valid for the optimization. If the lens is optimized without considering the incident light, the image quality of the peripheral region of the captured image is deteriorated during the experimental stage. Therefore, in the proposed optimization pipeline, the captured image is divided into several regions of interest (ROI), and a PSF that dominates each region is calculated. As a result, each region of the image is convolved with different PSFs and blurred in different ways. To keep the ROIs not overlapping, the location of ROIs are pre-determined by the number of total ROIs. By using specifications of image sensor and lens system, such as pixel pitch and resolution of the sensor and the distance between the lens and the sensor, incidence angles are determined. Figure 4.2 shows data pipeline for PSF calculation in the optimization and examples of PSF results

according to the group of Zernike coefficients. By using system configuration, which contains pretty much every optical parameter in Figure 4.1, the position and vector of point light sources are declared first. The light sources are converted to 4D light field information and propagates to the first lens surface which is the target surface of the optimization. The target surface is initialized with a certain curvature of lens or randomly generated group of coefficients. After initialization, the surface profile is calculated to compute the first refraction on the target surface. With vectorized Snell's law, the 4D light field is refracted and propagates to the backside of the lens. After the second refraction at the backside of the lens, it propagates to the plane of image sensor. On the plane where the image sensor is located, the 4D light field is integrated on angular domain and exploited as a source of raw PSF information. To gather the rays on the image sensor plane, the number of rays in a specific pixel needs to be counted. This process is basically a similar step of making histogram. When the library makes 2D histogram, each ray is considered as an individual particle which cannot be breakable. It means that most of computational library treat the index of input matrix to be converted to histograms as integer type of data. While making a histogram for raw PSF, therefore, the library internally converts data type to integer accordingly. Changing data type sometimes forces the pipeline non-differentiable. To avoid discontinuity, the process for making histogram is blocked and the auto-differentiation applied to input and output values for the block. The block is the only exception in the entire optimization pipeline, which is an inevitable decision to compose the pipeline.

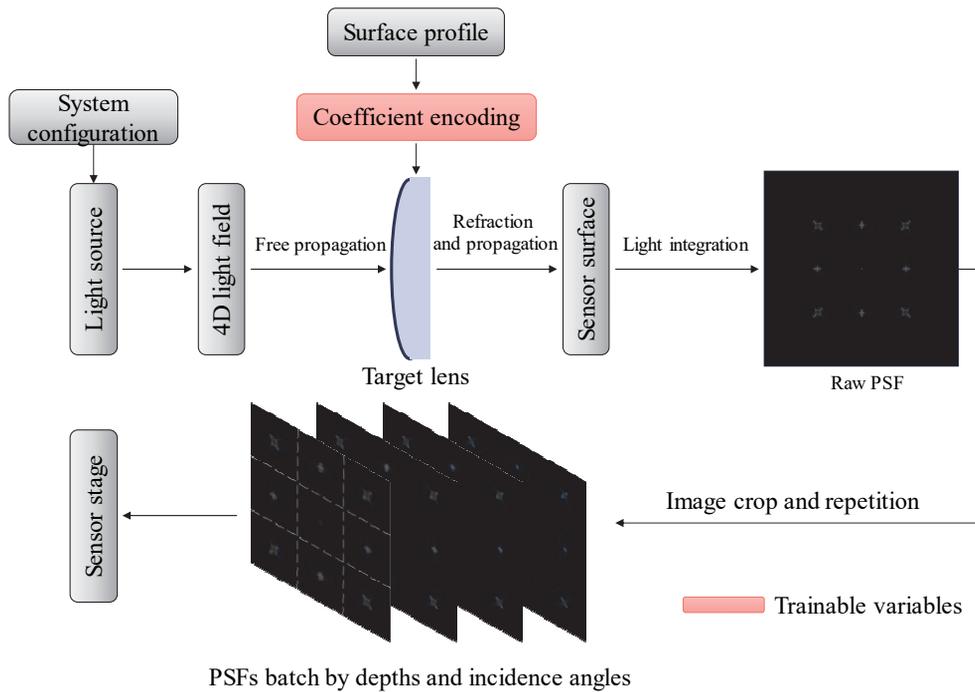


Figure 4.2 PSF calculation pipeline. Zernike coefficients are used to encode the target surface and declared as trainable variables.

4.2.2 Simulation for image acquisition at image sensor stage

Figure 4.3 shows the image acquisition on the image sensor including resizing, cropping and adding noise by the circuit including convolution computations. After calculations of PSFs for several ROIs, image processing for output image from the image sensor is followed. First, input images are blurred by the PSFs of lens in optimization process. To simulate the blur images, the PSFs are convolved with the input images according to ROIs and depths. Since the images are processed with different PSFs according to the location of ROI, the input images are sliced and centered to the allocated PSF. To avoid vignetting problems on each slice of image, the images are cropped with some margins or padded if there is no peripheral image. The convolution computation is implemented with Fourier and inverse Fourier transform while keeping overall procedures differentiable by the pipeline library. Following optical modulation by lens surface, the image sensor records the

blurred images by pixels. To simulate the image sensors, the specifications of the sensor are needed such as total number of pixels, distance between pixels, and wavelength characteristics of Bayer filter to determine wavelengths for each color channel. Depending on the area and resolution of the image sensor, the blurred images are cropped and resized to simulate imaging systems.

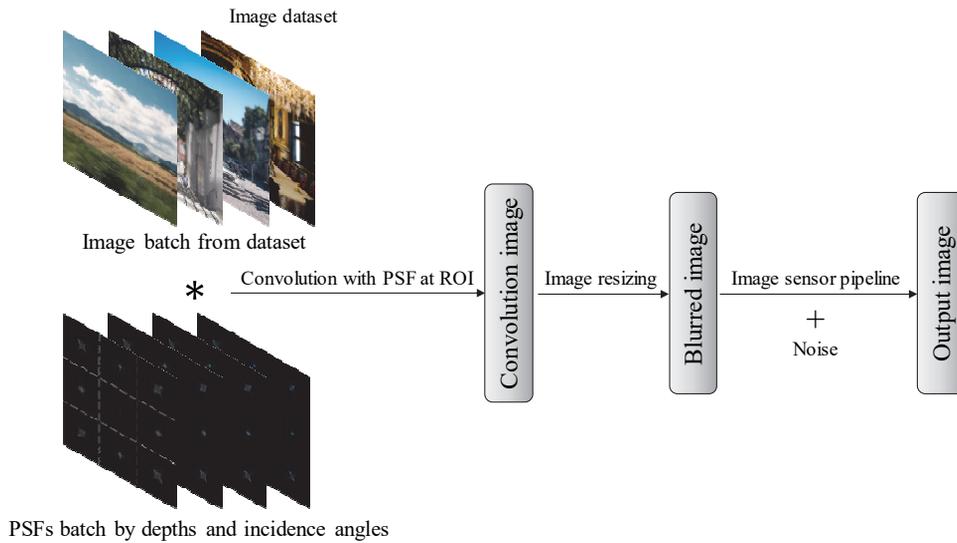


Figure 4.3 Image processing pipeline at the on sensor stage.

Cropping and resizing are determined by structural properties of the image sensor, which is related to physical size and numbers of pixels. While the light is integrated over time and converted to digital signal, electronic circuitry in the sensor makes noise on the output image. The read noise can be simulated with regard to camera parameters which can effect a noise level. In this chapter, it is supposed that the final noise on the image is normally distributed and added to input images after cropping and resizing. Assuming that the output data type of the actual image sensor is considered and that the image sensor has a proper exposure time so as to maintain the brightness of the input image, the brightness of images with noise is normalized and the total amount of information of the image is compressed if needed.

4.3 End-to-end optimization pipeline for designing optical element and image post-processing

4.3.1 Structure of end-to-end optimization pipeline

The deconvolved image with added noise becomes the output image of the image sensor stage. In this section, the output image and dataset images calculated in the above pipeline are compared to obtain the loss and update the parameters for the SGD algorithm. Specifically, the proposed ROI-based deconvolution method is introduced for consideration of off-axis light sources, and the loss calculation method used in this dissertation is proposed to implement the extended DOF.

Figure 4.4 shows a pipeline that takes the output image of the image sensor stage and reconstructs the image. In the PSF calculation process, the image batch used as input data is blurred by the PSFs calculated differently according to the ROI. The output image with noise becomes the final output of the image sensor stage. Output images are one of two inputs to the image post-processing pipeline. As shown in Figure 4.4, it is confirmed that the output image experiences different types of blur depending on the ROI. The other input to the image post-processing pipeline is the target PSF that will be used as the kernel to reconstruct the blurred image. Although the PSFs are obtained by depth using the light sources located in the target diopters in the PSF calculation stage, the pipeline uses only one PSF in the process of reconstructing the image. And the PSF can be considered as a target PSF. This procedure is a key assumption for implementing the extended DOF in this computational imaging study. A method of reconstructing an image through one-step of deconvolution for each ROI is implemented using a single PSF, regardless of an object placed at any depth. The PSF to be used as the kernel of deconvolution may be one of the PSFs of the target diopters or may be created independently. For example, suppose the user sets the target diopter to 0D, 1D, 2D, and 3D, and the imaging system wants the sharpest focus to have a depth of 2D. If

so, images blurred through the PSF created at each depth are reconstructed using only the PSF generated in 2D. Also, the target depth may be set differently according to the specification of the imaging system. It is possible that distant depths beyond the hyper focal distance need not be included in the target diopter, since the light starting from a distance farther than the hyperfocal distance of the imaging system is approximately the same as the hyperfocal distance PSF. Except for one selected depth, images are not completely reconstructed because different kernels are used for convolution and deconvolution. Therefore, the loss of the pipeline for applying the SGD algorithm is very large at first. As optimization proceeds, the lens surface is optimized to have a depth-invariant PSF to reduce system loss. However, even if an optical part has a freeform surface, the PSF made by the light source starting from different distances is very difficult to equal. Therefore, in this chapter, a method to learn a color kernel to be used for image reconstruction which is separate from the PSF generated from the light source of the target diopter is proposed. In addition, the images by deconvolution using the target depth PSF and independent kernel are compared and represented.

Various deconvolution algorithms can be used to reconstruct the output image of the sensor. In this dissertation, Wiener deconvolution, the simplest form of deconvolution algorithm, is used. Most image reconstruction algorithms use input images and a kernel for deconvolution as inputs. In addition, it has parameters for operating the algorithm. For example, Wiener deconvolution uses a parameter known to be most likely to reconstruct image best if it is proportional to the reciprocal of the signal-to-noise ratio between images and noise. The optimization pipeline also considers the parameters used for image reconstruction as a subject of optimization. As mentioned in the previous section, the noise added to the image reflects the characteristics of the sensor and its intensity and standard deviation are determined. However, in actual cases, it is difficult to know the characteristics of such noise in advance, and there is a possibility that it may change depending on

type of sensors. Therefore, setting the parameters of the reconstruction algorithm as the targets of the training is a way to cope with these various real situations.

To implement a deconvolution algorithm with ROIs, the images are cropped to fit the center and size of the ROI and the PSFs are processed as well. In order to prevent discontinuous image reconstruction from occurring at the boundary between the ROIs, the images at ROIs are cropped with a proper margin so as to be larger than the actual ROI size. For an ROI located at the edge of the image without information, a zero value is padded to where the margin is required. The PSF can be handled differently depending on the data structure of the PSFs. When a PSF is delivered in a batch format, the PSF is cropped at the center of each ROI including margin. When off-axis PSFs are added to an image to reduce the amount of data transmitted, as shown in Figure 4.3, each PSF is cropped along ROI position without any margin and then zero-padded to have the resolution of the cropped image. This is to prevent that the peripheral PSFs are cut together when cropping the PSF to the ROI including the margin. On the other hand, if the PSF is added to one color image plane, the distribution of each PSF may cross the border of ROIs boundary. This problem can be solved by initializing the surface to be optimized to have a certain level of curvature. Since the surface to be optimized acts as a lens fundamentally, this initialization has the effect of reducing the time it takes to optimize.

As shown in Figure 4.4, the reconstruction image is used to compute the loss of the system directly compared to the dataset images. To give flexibility to optimization, the image of the dataset was taken using a typical color camera and the pipeline doesn't perform any further normalization for preparing. However, in order to make the data necessary for optimization more random, it was cropped randomly and flipped left and right or up and down. Since the resolution of the image sensor should be the same as the resolution of the output image due to the system configuration information, the dataset image is resized so that the resolution

of the image sensor matches the resolution of the image sensor. Therefore, the resized dataset images are used as the ground truth and the L2-norm value of the difference from the reconstructed image is used as the loss of the system.

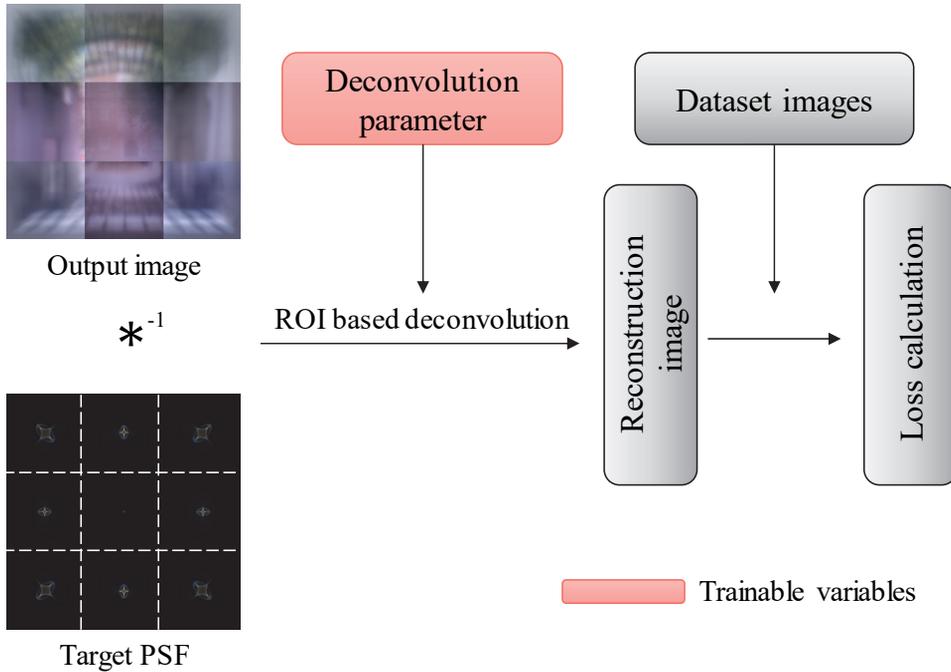


Figure 4.4 Image post-processing pipeline for calculating losses.

Figure 4.5 shows data flow of the proposed end-to-end optimization pipeline. The red-color boxes means that they are stages with training variables. As mentioned above paragraphs, PSFs are computed by using system configuration from the surface profile made by group of Zernike coefficients which is updated while optimization process. PSFs calculation stage makes a batch of PSFs according to the target depths and ROIs. The PSF batch is applied to the inputs of image sensor stage which calculate blurred images by the PSFs with image dataset. After resizing and cropping, image sensor stage makes output image which is ready for reconstruction in image post-processing pipeline. One of the PSF in the batch is used in image post-processing stage for the kernel of deconvolution.

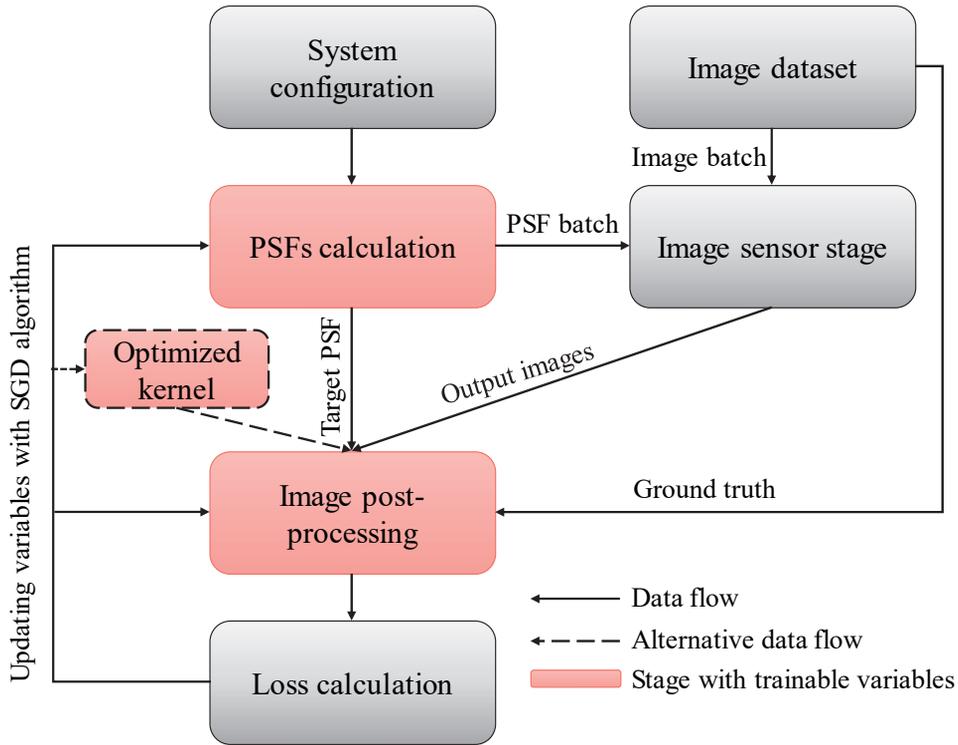


Figure 4.5 Data flow of the proposed end-to-end optimization pipeline. Trainable variables are updated by SGD algorithm.

However, using one of PSF as a target PSF for deconvolution makes the whole system having a small cost only for depth whose PSF is chosen as the target PSF. Therefore, the optimized system gives high quality of output images on the specific depth plane by victimizing image qualities of other depth plane. To design the extended DOF system having fair image quality through whole target depths, a method that using separate optimization kernel for deconvolution at image post-processing stage is proposed. The optimized kernel is also updated with SGD algorithm while the lens surface is optimized. As shown in the Figure 4.5 alternative data flow is represented with dash contour. The optimization of the reconstruction kernel can be done sequentially after optimizing the lens surface. However, in this case, the system performance is limited by the surface profile which has been already optimized. Optimizing deconvolution kernel and lens surface at the same time can be understood as a full search which is not realistic for

finding optimum results. To avoid the system fallen at local minimum, the optimized kernel should be initialized by one of the PSF of the initialized surface at the first training epoch.

As mentioned above paragraphs, the ground truth images are used to compute system loss by comparing reconstructed images. Various type of losses can be determined in the loss calculation stage. For example, to suppress divergences of the lens height, overall lens height can be treated as an element of the loss. Also, by using differential corrections on batch dimension, the reconstruction quality according to the target depths can be controlled. Since the learning rate for Zernike coefficients and optimized PSF should be extremely different, a constant is multiplied to learning rate of coefficients to optimize deconvolution kernel simultaneously.

4.3.2 Optimization of freeform lens and image processing for extended depth of field

Figure 4.6 shows a pilot simulation of proposed end-to-end optimization pipeline with a normal spherical surface. A BK-7 glass is used for the material of the lens. The wavelengths for color channels are 460 nm, 532nm, and 660 nm. Since the spherical surface cannot compensate color aberration at imaging surface, the curvature of the lens is designed to make the size of PSFs as small as possible. PSFs in the Figure 4.6 is magnified at the center of the images. The depths of the light sources are represented by diopter unit (D) which is a reciprocal of meters. The target depth is 1 D and its PSF information is represented with a red outline. The lens has a plano-convex shape and the curved surface is located towards objects. A pseudo-color mapping is used to represent colors of PSFs because the dynamic ranges of the PSFs cannot encoded by 8-bit images. The captured images show the blurred image on the image sensor plane after cropping and resizing according to the specifications of the image sensor. Group of captured images are reconstructed by deconvolving with the target PSF. Weiner deconvolution is used

to reconstruct images, which contains only one parameter. Reconstruction results in Figure 4.6 shows conventional spherical lens makes ringing noise on reconstructed images and is not appropriate for implementing extended DOF technique.

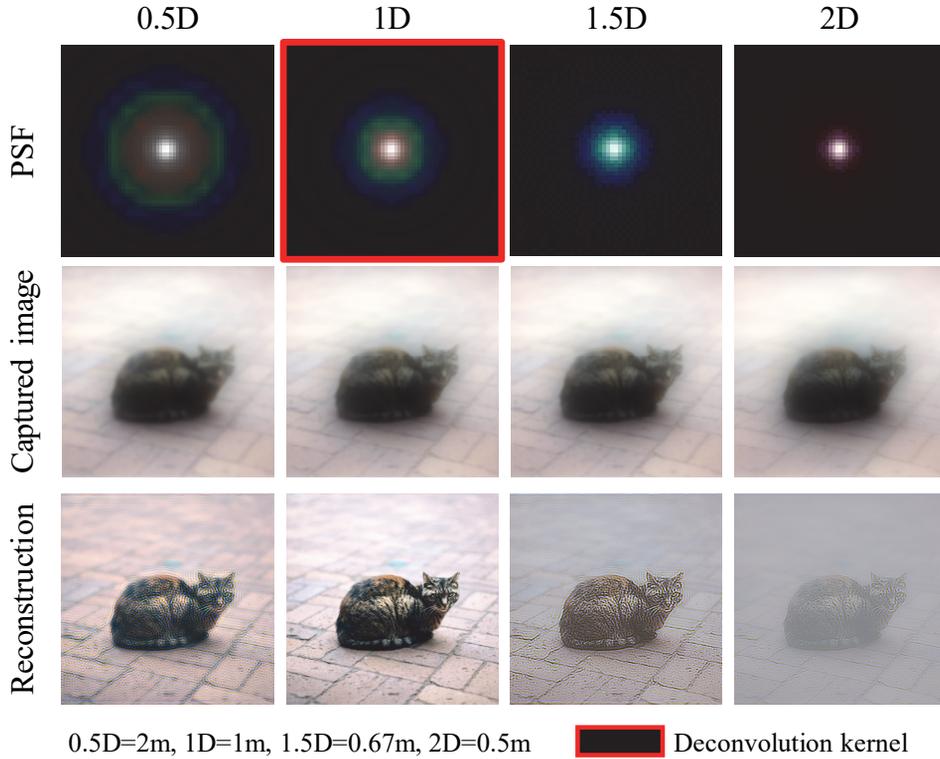


Figure 4.6 Spherical surface example for PSF calculation, image blurs and reconstructed images.

Figure 4.7 shows the simulation results of extended DOF by proposed end-to-end optimization pipeline. Target distances for extended DOF is 0.5 D to 2 D with a constant step on diopter unit. The PSF made by light on 1 D is used for deconvolution kernel. Compared to the shape of the spherical lens in Figure 4.6, the PSFs of the optimized lens have more detailed features around the center part of the PSFs. The initial value of the optimized lens is set to a normal spherical lens. The surface profile of the lens is encoded by first 21 Zernike coefficients and Adaptive moment estimation (ADAM) optimizer is used to implement SGD algorithm. Since the pipeline supports color channels, the implemented pipeline is

achromatic and showing stable results for various colors. Figure 4.7 represents two different images for showing the reconstruction results (*cat* and *balcony*). The *cat* results shows reconstruction performance for a range of various spatial frequency where the focused and defocused part of the images. The *balcony* represents the achromatic property of the system with colorful reconstruction results. By comparing the results of normal lens, the optimized result shows more natural and realistic images. Especially for the image from 2 D, the quality of the reconstruction images are improved as shown in the Figure 4.7. The ringing noise is also eliminated compared to the Figure 4.6. The diameter of the optimized lens is one inch and F-number of the system is 0.8, which is commercial level of numerical aperture. The resolution of the image sensor is 1200×1200 with $5.5 \mu\text{m}$ of pixel pitch. The crop ratio between a raw PSF and deconvolution kernel is 0.26, which is determined by the ratio of lens diameter and total size of the sensor. The number of ROIs is 3×3 and more detailed information is represented with a description of the Figure 4.8. Margin rate for seamless reconstruction between ROIs is 20 % of the each ROI region. By the amount of the computation load, the number of GPUs is controlled. A workstation with maximum two GPUs (Tesla P40, 24GB GDDR5 memory) is used for the pipeline.

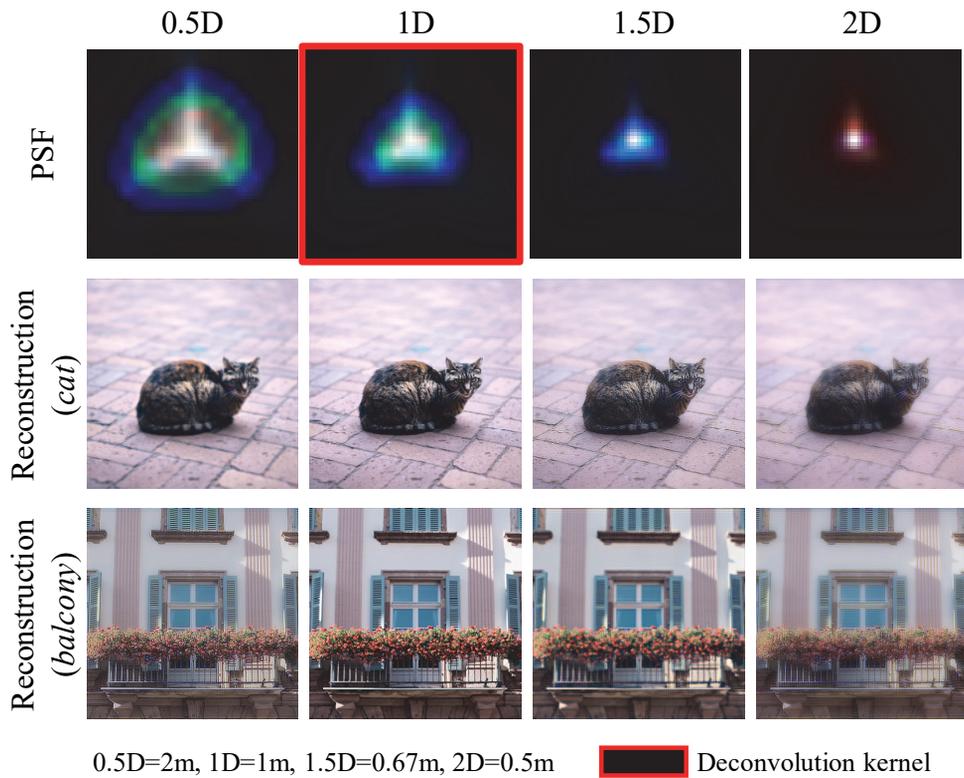


Figure 4.7 Simulation results of extended depth of field by using ray-based end-to-end optimization pipeline.

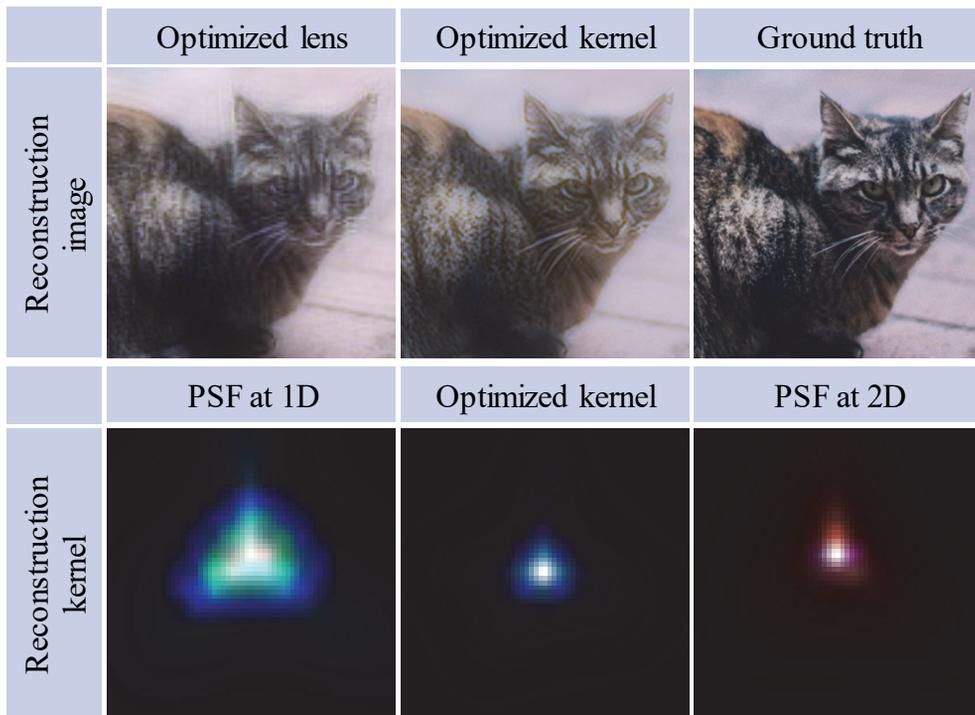


Figure 4.8 (upper row) Comparison of images from 2 D between ground truths, reconstruction results with 1 D PSF, reconstruction with optimized PSF and (lower row) PSF representations of 1 D, optimized kernel and 2 D.

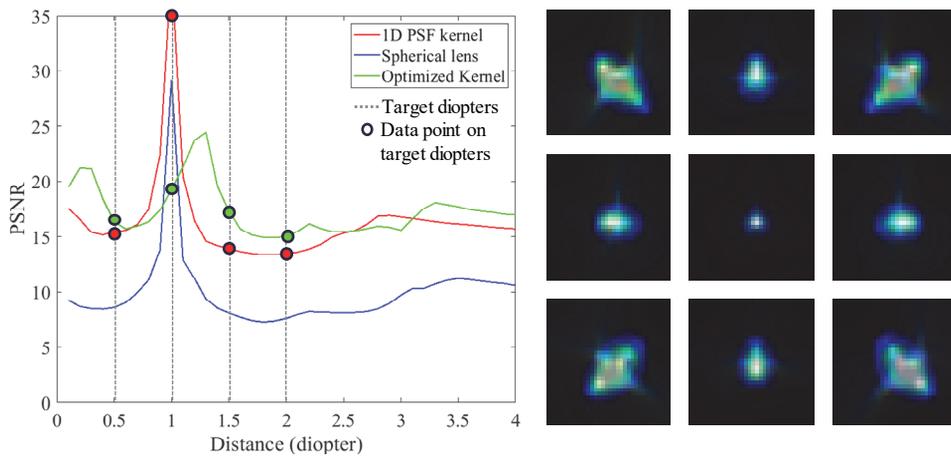


Figure 4.9 (left) PSNR plot according to the distance of subjects and (right) independently trained kernels for non-paraxial input light.

Figure 4.8 shows the results of using optimized kernel. The optimized kernel is a separate trainable variables in the pipeline. Since the usage of one of lens PSFs gives the best results at the chosen depth plane, the system performance is biased in the plane where the deconvolution kernel is made. For overall improvement of extended DOF performance, the independent kernel is optimized. The trainable variables for the kernel has 300×300 resolution and is padded with zeros to get the same resolution of the image sensor. Since the overall PSFs of the system has intensive information at the center of the image, the resolution of the kernel is intentionally limited to reduce computational load and learning time. As shown in Figure 4.8, the optimized kernel shows better results than usage of PSF on 2 D, which is represented in Figure 4.7 also. Because the target PSF is made by the light source at 1 D, the image on 2 D is represented to validate the simulation results, which is the farthest from the target in diopter unit. Figure 4.9 shows PSNR graphs of the reconstruction imaging by comparing the ground truths. The gray lines represent the target distance of the optimization pipeline. For red and blue line, PSFs made from 1 D are used for reconstruction and the green line shows the results of using independent deconvolution kernel. As shown in the graph, using independent kernel gives the stable results according to the distances. Compared to the independent kernel version, using the real PSF makes the best results at the original plane of light source for deconvolution PSF. The right figure on Figure 4.9 shows the independently trained kernels for non-paraxial light source. Nine RoIs are considered to represent non-paraxial illumination. Each PSF shows different shapes and aberration which make the quality of peripheral image improved.

Figure 4.10 shows PSNR plot according to target depth ranges. Independent kernel method is used to optimize the pipeline. In this simulation, three target distances are used which is centered 1.5 D. The performance for extended depth of field increase at narrow target depth range.

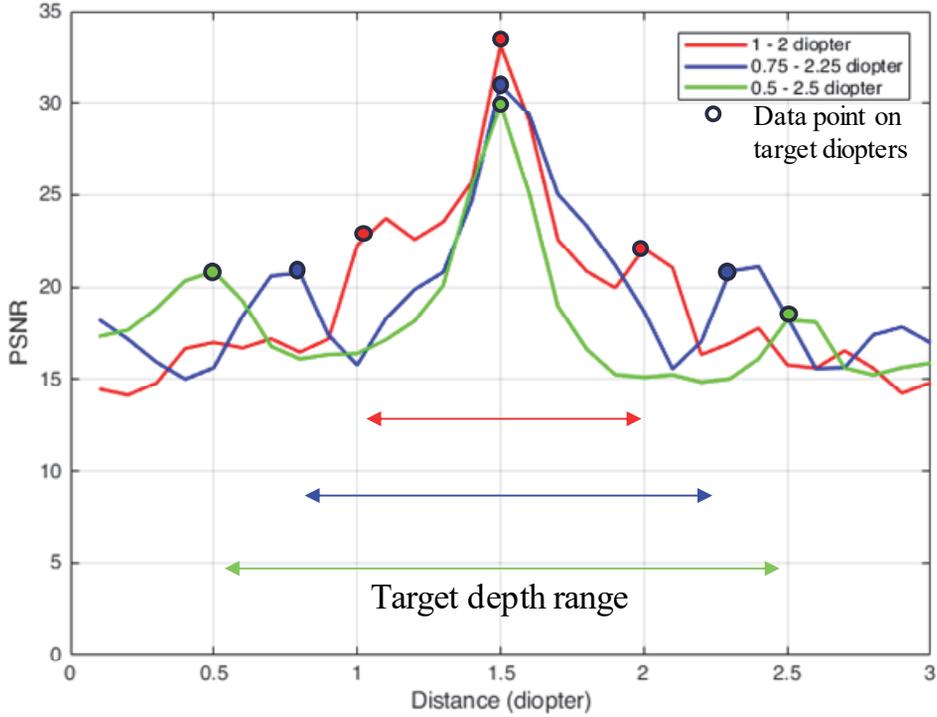


Figure 4.10 PSNR plot according to target depths range

4.4 Conclusion

In this chapter, ray-based end-to-end optimization pipeline for designing optics and image post-processing is proposed. To overcome the limitations on the curvature of the lens surface in wave-based optimization pipeline, the ray-tracing method is exploited to calculate the PSF of the target surface. The pipeline is implemented as a fully-differentiable processing to adopt auto-differentiation of the deep learning library. The purpose of the computational imaging is extended depth of field and a domain-specific lens surface and a deconvolution method are optimized simultaneously. The optimized surface is encoded with Zernike coefficients which are the trainable variables of the pipeline. Also parameters in deconvolution algorithm are set to be a trainable variables. By optimizing the optical element and image post-processing algorithm, the proposed optimization method can achieve improvements on image quality of reconstruction results. Additionally, ray-based

PSF calculation method supports commercialization-level of numerical aperture and the curvature of the optimized lens. Also, the light-field information of the light source is used for considering non-paraxial light. For the extended depth of field, two types of kernels for deconvolution were introduced: PSF of target distance and optimized kernel. By using independent optimized kernel, more stable reconstruction results can be generated along the various target depths. The target surface had a half inch diameter and viewing angle of the system was larger than 30 degree. The F-number of the imaging system is less than 1, which is similar to that of a commercially available lens. The performance of the extended depth of field is analyzed according to the kernel types for deconvolution algorithm.

Chapter 5. Conclusion

There is a growing demand for imaging systems that utilize the latest technologies and support multi-functionality. In this dissertation, several image processing techniques are studied for light-field based computational imaging system, including LFCs and end-to-end optimized imaging. Unlike previous research on LFCs, which was devoted to improving image post-processing technology, LFCs was extensively researched based on its optical configuration and specifications. Also, image processing methods for LFCs were introduced, which is capable of analyzing light-field information according to the several optical and camera parameters.

Besides, an algorithm for extracting depth information of a captured object adaptively and accurately was proposed. The image processing method reflecting the optical configuration of the LFC had the advantage that the depth information of the photographed image can be precisely extracted in real distance units. The extracted depth information had a real distance unit, and 3D design software and autostereoscopic displays reconstructed the subjects. The proposed LFC and image processing technology was proved to be suitable for obtaining the elemental image of the next-generation display.

Also, researches were carried out to design and optimize computational imaging technology using deep learning software library, including image processing for light-field information and considering the non-paraxial light source. Conventional computational imaging techniques were developed in terms of image processing independently or using heuristic cost function for freeform optics. In this dissertation, a technique was proposed to simultaneously optimize optical elements and image restoration methods suitable for computational imaging using ray-tracing based technology. The entire pipeline from input to reconstruction of the image was implemented to be fully differentiable. The designed device enabled

extended depth of field through computational imaging and included the design for off-axis input light so that the proposed optimization technique can be used in optical design with larger numerical apertures.

The proposed image processing technique for LFC and computational imaging in this dissertation is expected to be used as future imaging technology and also work as a camera capable of acquiring information for the next-generation display.

Bibliography

1. E. H. Adelson and R. B. James, "The plenoptic function and the elements of early vision," 3-20 (1991).
2. J. W. Goodman, Introduction to Fourier optics, Roberts and Company Publishers (2005).
3. G. Lippmann, "La photographie integrale," Comptes-Rendus, Académie des Sciences, vol. 146, 446-451 (1908).
4. R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a handheld plenoptic camera," Computer Science Technical Report CSTR, vol. 2, no. 11, 1-11 (2005).
5. V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," ACM Transactions on Graphics (TOG), vol. 37, no. 4, 114 (2018).
6. F. Heide, Q. Fu, Y. Peng, and W. Heidrich, "Encoded diffractive optics for full-spectrum computational imaging," Scientific reports, vol. 6, 33543 (2016).
7. 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," Applied Optics, vol. 50, no. 34, H87-H115 (2011).
8. J. Y. Son, B. Javidi, S. Yano, and K. H. Choi, "Recent developments in 3-D imaging technologies," Journal of Display Technology, vol. 6, no. 10, 394-403 (2010).
9. S.-g. Park, J.-Y. Hong, C.-K. Lee, M. Miranda, Y. Kim, and B. Lee, "Depth-expression characteristics of multi-projection 3D display systems," Applied optics, vol. 53, no.27, G198-208 (2014).
10. B. Lee, "Three-dimensional displays, past and present," Physics Today, vol. 66, no. 4, 36-41 (2013).
11. J.-H. Park, K. Hong, and B. Lee, "Recent progress in three-dimensional information processing based on integral imaging," Applied Optics 48, H77-H94 (2009).

12. B. Javidi and F. Okano, eds., *Three Dimensional Television, Video, and Display Technology*, Springer Science & Business Media (2002).
13. 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, 1598-1603 (1997).
14. S. K. Lee, S. I. Hong, Y. S. Kim, H. G. Lim, N. Y. Jo, and J. H. Park, "Hologram synthesis of three-dimensional real objects using portable integral imaging camera," *Optics Express*, vol. 21, no. 20, 23662-23670 (2013).
15. K. Hong, J. Hong, J.-H. Jung, J.-H. Park, and B. Lee, "Rectification of elemental image set and extraction of lens lattice by projective image transformation in integral imaging," *Optics Express*, vol. 18, no. 11, 12002-12016 (2010).
16. H. Navarro, R. Martínez-Cuenca, G. Saavedra, M. Martínez-Corral, and B. Javidi, "3D integral imaging display by smart pseudoscopic-to-orthoscopic conversion (SPOC)," *Optics Express*, vol. 18, no. 25, 25573-25583 (2010).
17. J.-H. Jung, J. Kim, and B. Lee, "Solution of pseudoscopic problem in integral imaging for real-time processing," *Optics Letter*, vol. 38, no. 1, 76-78 (2013).
18. J. Arai, F. Okano, H. Hoshino, and I. Yuyama, "Gradient-index lens-array method based on real-time integral photography for three-dimensional images," *Applied Optics*, vol. 37, no. 11, 2034-2045 (1998).
19. J. Kim, J.-H. Jung, C. Jang, and B. Lee, "Real-time capturing and 3D visualization method based on integral imaging," *Optics Express* 21, no. 16, 18742-18753 (2013).
20. G. Li, K.-C. Kwon, G.-H. Shin, J.-S. Jeong, K.-H. Yoo, and N. Kim, "Simplified integral imaging pickup method for real objects using a depth camera," *Journal of the Optical Society of Korea*, vol. 16, no. 4, 381-385 (2012).
21. J.-S. Jang and B. Javidi, "Formation of orthoscopic three-dimensional real images in direct pickup one-step integral imaging," *Optical Engineering*, vol. 42, no. 7, 1869 (2003).
22. F. Okano, J. Arai, K. Mitani, and M. Okui, "Real-time integral imaging based on extremely high resolution video system," *Proceedings of the IEEE*, vol. 94, no. 3, 122-125 (2006).

23. J. Hahn, Y. Kim, E.-H. Kim, and B. Lee, "Undistorted pickup method of both virtual and real objects for integral imaging," *Optics Express*, vol. 16, no. 18, 13969-13978 (2008).
24. J. Arai, F. Okano, M. Kawakita, M. Okui, Y.i Haino, M. Yoshimura, M. Furuya, and M. Sato, "Integral Three-Dimensional Television Using a 33-Megapixel Imaging System," *Journal of Display Technology*, vol. 6, no. 10, 422 (2010).
25. J. Arai, H. Kawai, and F. Okano, "Microlens arrays for integral imaging system," *Applied Optics*, vol. 45, no. 36, 9066–9078 (2006).
26. A. Lumsdaine and G. Todor, "Full resolution lightfield rendering," Indiana University and Adobe Systems, Adobe Technical Report, vol. 92, 92 (2008).
27. J. Kim, J.-H. Jung, Y. Jeong, K. Hong, and B. Lee, "Real-time integral imaging system for light field microscopy," *Optics Express*, vol. 22, no. 9, 10210-10220 (2014).
28. M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, 924-934 (2006).
29. A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *Proceedings of International Conference on Computational Photography (ICCP)*, 1-8 (2009).
30. T. Bishop and P. Favaro, "The light field camera: extended depth of field, aliasing and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, 972-986 (2012).
31. J. Yeom, K. Hong, Y. Jeong, C. Jang, and B. Lee, "Solution for pseudoscopic problem in integral imaging using phase-conjugated reconstruction of lens-array holographic optical elements," *Optics Express*, vol. 22, no. 11, 13659-13670 (2014).
32. K. Yamamoto, T. Mishina, R. Oi, T. Senoh, and M. Okui, "Cross talk elimination using an aperture for recording elemental images of integral photography," *Journal of Optics Society of America A*, vol. 26, no. 3, 680–690 (2009).

33. F. Jin, J. S. Jang, and B. Javidi, "Effects of device resolution on three-dimensional integral imaging," *Optics Letter*, vol. 29, no. 12, 1345-1347 (2004).
34. 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, L71-L74 (2005).
35. J. H. Park, S. K. Lee, N. Y. Jo, H. J. Kim, Y. S. Kim, and H. G. Lim, "Light ray field capture using focal plane sweeping and its optical reconstruction using 3D displays," *Optics Express*, vol. 22, no. 21, 25444-25454 (2014).
36. J. R. Parker, *Algorithms for image processing and computer vision* (John Wiley & Sons, 2010).
37. K.-C. Kwon, C. Park, M.-U. Erdenebat, J.-S. Jeong, J.-H. Choi, N. Kim, J.-H. Park, Y.-T. Lim, and K.-H. Yoo, "High speed image space parallel processing for computer-generated integral imaging system," *Optics Express*, vol. 20, no. 2, 732-740 (2012).
38. F.C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 60, 1–12 (2015).
39. D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, 220 (2013).
40. H. Hua and B. Javidi, "A 3D integral imaging optical see-through head-mounted display," *Optics Express*, vol. 22, no. 11, 13484–13491 (2014).
41. E. Moon, M. Kim, J. Roh, H. Kim, and J. Hahn, "Holographic head-mounted display with RGB light emitting diode light source," *Optics Express*, vol. 22, no. 6, 6526–6534 (2014).
42. Y. Takaki, Y. Urano, S. Kashiwada, H. Ando, and K. Nakamura, "Super multi-view windshield display for long-distance image information presentation," *Optics Express*, vol. 19, no. 2, 704–716 (2011).
43. J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern recognition*, vol. 37, no. 4, 827–849 (2004).
44. J. Geng, "Structured-light 3D surface imaging: a tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, 128–160 (2011).

45. J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images." *Communications of the ACM*, vol. 56, no. 1, 116–124 (2013).
46. J. Zhu, L. Wang, R. Yang, and J.E. Davis, "Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 7, 1400–1414 (2011).
47. S.K. Lee, S.I. Hong, Y.S. Kim, H.G. Lim, N.Y. Jo, and J.H. Park, "Hologram synthesis of three-dimensional real objects using portable integral imaging camera," *Optics Express*, vol. 21, no. 20, 23662–23670 (2013).
48. R. Prevedel, Y.G. Yoon, M. Hoffmann, N. Pak, G. Wetzstein, S. Kato, T. Schrödel, R. Raskar, M. Zimmer, E.S. Boyden, and A. Vaziri, "Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy," *Nature methods*, vol. 11, no. 7, 727–730 (2014).
49. C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch, "Refocusing distance of a standard plenoptic camera," *Optics Express*, vol. 24, no. 19, 21521–21540 (2016).
50. Y. Jeong, J. Kim, J. Yeom, C.K. Lee, and B. Lee, "Real-time depth controllable integral imaging pickup and reconstruction method with a light field camera," *Applied Optics*, vol. 54, no. 35, 10333–10341 (2015).
51. M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," *Optics Express*, vol. 13, no. 23, 9175–9180 (2005).
52. C.M. Do and B. Javidi, "3D integral imaging reconstruction of occluded objects using independent component analysis-based K-means clustering," *Journal of display technology*, vol. 6, no. 7, 257–262 (2010).
53. H.G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1547–1555 (2015).
54. F. Liu, G. Hou, Z. Sun, and T. Tan, "High quality depth map estimation of object surface from light-field images," *Neurocomputing*, 252, 3–16 (2017).

55. D.G. Dansereau, I. Mahon, O. Pizarro, and S.B. Williams, "Plenoptic flow: Closed-form visual odometry for light field cameras," In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, 4455–4462 (2011).
56. E. Iffa, G. Wetzstein, and W. Heidrich, "Light field optical flow for refractive surface reconstruction," in Applications of Digital Image Processing XXXV, vol. 8499, 84992H (2012).
57. J.H. Jung, K. Hong, G. Park, I. Chung, J.H. Park, and B. Lee, "Reconstruction of three-dimensional occluded object using optical flow and triangular mesh reconstruction in integral imaging," Optics Express, vol. 18, no. 25, 26373–26387 (2010).
58. A. Metallo, V. Rossi, J. Blundell, G. Waibel, P. Graham, G. Fyffe, X. Yu, and P. Debevec, "Scanning and printing a 3D portrait of president Barack Obama," In Proceedings of SIGGRAPH 2015: Studio, 09–13 (2015).
59. H. Todoroki and H. Saito, "Light field rendering with omni-directional camera," in Visual Communications and Image Processing, vol. 5150, 1159–1169 (2003).
60. B.K. Horn and B.G. Schunck, "Determining optical flow," Artificial intelligence, vol. 17, no. 1-3, 185–203 (1981).
61. S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," International Journal of Computer Vision, vol. 92, no. 1, 1–31. (2011).
62. G. Passalis, N. Sgouros, S. Athineos, and T. Theoharis, "Enhanced reconstruction of 3D shape and texture from integral photography images," Applied Optics, vol. 46, 5311–5320 (2007).
63. H. Kim, J. Hahn, and B. Lee, "Mathematical modeling of triangle-mesh-modeled three-dimensional surface objects for digital holography," Applied Optics, vol. 47, D117–D127 (2008).
64. J. Jeong, J. Cho, C. Jang, G. Li, and B. Lee, "Simple quality improvement method for holographic display using digital micro-mirror device," In Proceedings of Imaging and Applied Optics, JW4A–43 (2016).

65. H. Kim, C.Y. Hwang, K.S. Kim, J. Roh, W. Moon, S. Kim, B.-R. Lee, S. Oh, and J. Hahn, "Anamorphic optical transformation of an amplitude spatial light modulator to a complex spatial light modulator with square pixels," *Applied Optics*, vol. 53, no.27, G139–G146 (2014).
66. Y. Peng, Q. Fu, F. Heide, and W. Heidrich, "The diffractive achromat full spectrum computational imaging with diffractive optics," In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, 4 (2016).
67. L.L. Carter and E.D. Cashwell, *Particle-transport simulation with the Monte Carlo method*, Los Alamos Scientific Lab (1975).

Appendix

Portions of the work discussed in this dissertation were also presented in the following publications:

[Chapter 2] Y. Jeong, J. Kim, J. Yeom, C.K. Lee, and B. Lee, "Real-time depth controllable integral imaging pickup and reconstruction method with a light field camera," *Applied Optics*, vol. 54, no. 35, 10333–10341 (2015)

[Chapter 3] Y. Jeong, S. Moon, J. Jeong, G. Li, J. Cho, and B. Lee, "One shot 360-degree light field capture and reconstruction with depth extraction based on optical flow for light field camera," *Applied Sciences*, vol. 8, no. (6), 890 (2018).

[Chapter 4] Y. Jeong, J. Cho, D. Yoo, and B. Lee, "Optic design and image processing considering angle of incidence via end-to-end optimization method," In *Ultra-High-Definition Imaging Systems II*, vol. 10943, 109430U (2019).

초 록

광학 기기와 카메라가 포함된 모바일 기기들의 거대한 상업적 성공으로 인해 단순 촬영 이상의 기능을 지원하는 차세대 전산 카메라에 대한 연구가 진행되고 있다. 이미지 센서의 소형화와 이미지 프로세서 성능의 향상은 그러한 연구 및 상용화의 연착륙을 이끌어 냈다. 또한 이로 말미암아, 최신 기술을 활용하는 다기능 이미징 시스템에 대한 수요가 최근 증가하고 있다.

본 논문에서는 광학적 구성을 이미징 시스템 설계 및 복원 방법에 고려한 전산 영상 시스템과 이를 위한 이미지 처리 방법에 대한 연구가 소개된다. 라이트-필드 카메라는 깊이와 색상 정보를 동시에 기록하여 기존 모바일 카메라를 대체 할 수 있는 대안으로 제안되었다. 그러나, 라이트-필드 카메라에 대한 기존의 연구는 이미지 후 처리 기술을 독립적으로 향상시키고 카메라의 광학적 특성을 고려하지 않는 경향이 있다. 라이트-필드 카메라와 촬영된 이미지의 복원 방법을 광학적 특성에 맞추어 설계하기 위해, 광학 요소들의 배치와 라이트-필드 카메라의 시스템 특성이 카메라의 디자인과 복원 방법을 설계하는데 반드시 참조해야 한다.

이와 같은 기초아래, 정확한 F값 매칭 조건, 광학 구성에 대한 적응형 깊이 추출 방법을 포함하여 라이트-필드 카메라의 설계 방법들이 연구되었다. 본 논문에서 라이트-필드 카메라에 대한 연구는 시스템의 사양을 고려하기 때문에 이미지 후 처리에서 추출된 모든 정보는 실제 거리 단위를 갖는다. 따라서 깊이 맵 영상과 복원된 3차원 모델은 증강현실과 가상현실에 이용 가능한 차세대 무안경식 3차원 디스플레이의 요

소 영상으로 변환될 수 있다.

또한, 딥-러닝 소프트웨어 라이브러리를 사용하여 더욱 일반적인 형태의 전산 영상 시스템을 설계하고 최적화하기 위한 연구가 수행되었다. 종래의 전산 이미징 기술은 영상 처리를 독립적으로 개발하거나 자유형 광학을 위한 휴리스틱 (heuristic) 비용 함수를 사용하는 한계점이 있었다. 본 논문에서는 광선 추적 기술을 이용한 전산 이미징의 목적에 적합한 광학 소자와 영상 복원 방법을 동시에 최적화하는 기술이 제안된다. 나아가 비-근축 빛들이 복원 이미지의 주변부 화질을 향상 시키기 위해 고려된다. 이미지 입력에서 이미지 재구성까지의 전체 파이프 라인 은 미분 가능하도록 구현되어 목적으로 하는 표면과 이미지 복원 방법을 확률적 기울기 강하 (stochastic gradient descent) 기술을 통해 최적화 할 수 있도록 구현된다. 제안된 시스템은 전산 이미징을 통해 확장 된 피사계 심도 기술을 가능하게 하고 제안된 최적화 기술을 더 큰 수치 개구를 갖는 광학 설계에 사용할 수 있도록 축외 (off-axis) 입력 광에 대한 설계를 포함한다.

본 논문에서 제안한 라이트-필드 기반의 전산 이미징 기술을 위한 다양한 이미지 처리 방법들은 미래의 이미징 기술에 이용될 것으로 예상되며, 차세대 디스플레이를 위한 정보 습득을 할 수 있는 카메라 시스템 과 함께 이용될 것으로 기대된다.

주요어 : 라이트필드 카메라, 전산 이미징, 이미지 처리, 증강 현실, 렌즈 설계, 자유 곡면 소자

학 번 : 2013-20878