

國立交通大學

光電工程學系碩士班

碩士論文

利用變焦影像產生高動態深度範圍之深度圖

High Dynamic Depth Range Depth Map Rendering

by Adjusted Focused Images

研究生： 蘇雍仁

指導教授： 謝漢萍 博士

黃乙白 博士

中華民國一百零二年八月

利用變焦影像產生高動態深度範圍之深度圖

High Dynamic Depth Range Depth Map Rendering

by Adjusted Focused Images

研究生： 蘇雍仁

Student： Yong-Ren Su

指導教授： 謝漢萍

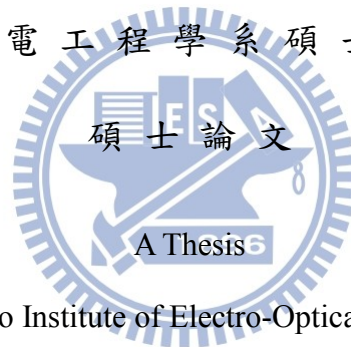
Advisors： Han-Ping David Shieh

黃乙白

Yi-Pai Huang

國立交通大學

光電工程學系碩士班



Submitted to Institute of Electro-Optical Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Electro-Optical Engineering

August 2013

Hsinchu, Taiwan, Republic of China

中華民國一百零二年八月

利用變焦影像產生高動態深度範圍之深度圖

碩士研究生： 蘇雍仁

指導教授： 謝漢萍 教授

黃乙白 教授

國立交通大學電機學院

光電工程學系碩士班

摘 要

隨著立體形象化技術的進步，在日常生活中，不同的娛樂方式如雨後春筍般蓬勃發展。人們可以看立體電影或與遊戲中的物件做三維互動，這的確是近年來全新的體驗。科學家與研究者也一直希望能發展一套系統可以提供人們有更真實的感受，然而無可避免地，我們渴望能夠顯示更豐富的內容，只是間接獲取深度資訊的技術時常受限於影像的內容，對消費者來說，利用相機陣列拍照也是不切實際的。所以我們選擇了實用的單相機系統，並且也利用透鏡陣列來擷取具有視差的基本影像。除此之外，我們之所以利用深度圖來表示我們的三維資訊是因為深度圖可以被應用於一般二維多工式的三維顯示器的立體影像組合成。但能夠分析視差的前提是基本影像需要是全聚焦的。對一般攝影來說，調整光圈來延展景深是最直接的方式，然而光圈太小時，當場景是昏暗的以及有移動的物體時，會碰到曝光時間的兩難。所以我們承繼了高動態範圍影像的概念，發展了一套高動態深度範圍的系統來產生一張具有廣闊工作範圍的深度圖。我們所提出的高動態深度範圍系統不僅提供了較我們相機裡最小光圈(150cm)更廣的深度範圍(165cm)，並且也縮短了至少 21 倍的曝光時間。而這樣的特性可以使得我們可以擷取快速移動的物體，卻不受一般相機大光圈的淺景深所限制。

High Dynamic Depth Range Depth Map Rendering by Adjusted Focused Images

Student: Yong-Ren Su **Advisors:** Dr. Han-Ping David Shieh
Dr. Yi-Pai Huang

**Institute of Electro-Optical Engineering
National Chiao Tung University**

Abstract

With the improvement of 3D visualization, many recreational activities spring up in our daily life. People can watch 3D movies or interact with 3D video games. It is indeed a brand-new experience around these years. A lot of scientists and researchers keep working on developing a system that can provide humans with more realistic perception. To accomplish this great task, longing for more abundant 3D content is inevitable. However, indirect rendering of depth information is often limited by the image categories, and it is impractical for customers to use camera array to create their own 3D content. As a result, single camera system is chosen for its practicability and we also utilize lens array to generate disparity in the elemental images. Moreover, we represent our 3D information by a depth map because it can be used to synthesize stereo image pairs for conventional multiplexed-2D 3D displays. Nonetheless, the prerequisite of stereo matching is all-in-focus elemental image. Increasing f-number to elongate depth of field is the most common way in photography, but this strategy confronts the dilemma of exposure time when the scene is dim and contains moving objects. Hence, we propose a high dynamic depth range (HDDR) system to render a depth map with wide working range, and this idea stems from the concept of high dynamic range (HDR)

images. Our proposed HDDR system not only provides wider working range (165 cm) than that of the largest f-number in our camera (150 cm), but also minimizes the exposure time by at least 21 times. And this characteristic will make it possible to capturing the object with quick movement but without limitation of the shallow depth of field of small f-number.



誌 謝

如果不回顧過去幾個春夏秋冬，我不會相信此時此刻我將要置身於另一段人生了；如果沒有大家一直以來的陪伴與幫忙，也許我已經迷路走散、在街上流浪了。更幸運的是，我有兩個風格截然不同的指導教授教導我。謝漢萍老師就像藏身武林的大俠一樣，總能在研究上，精闢地點出我所欠缺的部份；黃乙白老師則像是大哥一樣，除了學識上的傳授之外，心情上的難題也幫了不少忙。兩位老師對我的付出，我由衷感謝，除此之外，也很感激他們提供了豐沛的資源，與一個溫暖的實驗室。然而完成一篇論文，還有不能忘的口試委員：戴亞翔、歐陽盟、陳政寰老師，感謝他們最後的醍醐灌頂。

有人說做研究是寂寞的，但這句話在我們實驗室裡是不可能成立的，原因是我們盡心盡力的博士班學長姐們，特別是從大學部專題開始就一直帶著我的致維哥，這個稱呼代表了他的親切，也代表了我對他的崇拜。雖然要感謝他的事情細數不盡，但有幾樣是令我最感動的：包容與鼓勵。他願意去了解每一個不同的想法，也包容每一次實驗的失誤，最後都還能鼓勵我，陪著我將每項事情完成。身教重於言教，他做到了。直到後來換了題目後，開始與博六學長合作，雖然一開始兩個人都還在磨合，不過最後漸漸有了默契，還會懷念一起做實驗到半夜的日子，此外，也感謝他給我相當大的自主權，可以調配自己的時間，決定自己的進度。最強大的3D液晶組，還有兩位元老：頭哥和台翔。這兩位學長也是常被我騷擾的苦主，感謝他們實驗上的幫忙以及聆聽我的牢騷。除了博士班之外，還要感謝一個幾乎像我的哆啦A夢的學長，傳說中的光電王子，從我還是小大一的時候就受他照顧到念研究所，印象最深刻的還是待在無塵室裡，偶而被你臭罵的日子。

可不可以說：「我親愛的同學們，如果你們不是酒肉朋友，為什麼我腦袋裡都是玩樂的記憶？」人很好會載我的曜曜、只有說要一起減肥卻沒一起挨餓的小黑、泛舟想嚇死我之在我面前掉下去的小岡、脊椎側彎而免役的不MAN男米克、屁話無窮無盡的阿昌、以及說是田徑隊卻常昏倒的小靖，你們都是我最棒的夥伴。

不勝枚舉的實驗室同仁，你們的好我當然都記得，但若只列出名字卻反而讓我對你們的感謝顯得隨便，於是乎，你們組成了無與倫比的FPD&ADO LAB，最後也將在我心中組成了這段在學校裡不可磨滅的回憶。

最後，雖然帶不走交大這棵大樹，卻可以帶走交大精神——「知新致遠、崇實篤行」。

Contents

摘 要	i
Abstract	ii
誌 謝	iv
Contents	v
Figure Captions.....	vii
List of Tables	xii
Chapter 1 Introduction	1
1.1 Preface	1
1.2 Prior Researches	4
1.2.1 2D-3D Conversion.....	4
1.2.2 Multi-camera System.....	7
1.2.3 Sweep Focus.....	10
1.2.4 Light Field Camera.....	13
1.2.5 Integral Image.....	18
1.3 Motivation and Objective	24
1.4 Organization of the thesis	26
Chapter 2 Theory	27
2.1 Principle of 3D images	27
2.2 Microscopy	31
2.2.1 Optical Terminology	31
2.2.2 Optical Microscope.....	35
2.3 High Dynamic Range Imaging.....	39
2.3.1 High Dynamic Range Imaging Rendering Algorithm.....	40
Chapter 3 Structure and Algorithms.....	42
3.1 3D Image Capturing with Lens Array	42
3.2 Algorithm	49
3.2.1 Depth Estimation Reference Software (DERS).....	50
3.2.2 Depth Map Fusion from Edge Exploring Thresholding (DFEET).....	53

Chapter 4	<i>Experiments and Results</i>	58
4.1	Depth Maps of Under-exposed and Blurred Images	58
4.2	HDDR Depth Map Rendering of Two Depth of Field	62
4.3	HDDR Depth Map Rendering of Three Depth of Field	66
4.4	From Temporal to Spatial HDDR System.....	73
4.4.1	Temporal HDDR System	74
4.4.2	Spatial HDDR System	76
4.5	Discussion.....	78
Chapter 5	<i>Summary</i>	80
5.1	Conclusion	80
5.2	Future Work.....	82
5.2.1	Liquid Crystal Lens Array	83
5.2.2	Fine Depth Resolution	84
Reference	86

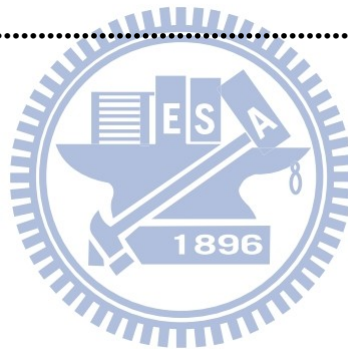


Figure Captions

Figure 1-1	Good old days	1
Figure 1-2	The style of dilettante	2
Figure 1-3	The evolution of displays	2
Figure 1-4	Prospect of virtual reality	3
Figure 1-5	Categorization of 3D capturing techniques	4
Figure 1-6	Principle of MTD	5
Figure 1-7	Generation of depth map by fusion	6
Figure 1-8	Scheme of concentric mosaics method, (a) top view (b) side view	7
Figure 1-9	Setup of longitudinal aligned camera array system	8
Figure 1-10	Setup to reconstruct the entire shape	9
Figure 1-11	Focal stack or space-time focal volume	10
Figure 1-12	Depth of field of multiple shots	10
Figure 1-13	Chromatic dispersion of focal lengths	11
Figure 1-14	Parameterization of light slab	13
Figure 1-15	Conceptual scheme of light field camera	14
Figure 1-16	Parameters of synthetic photography equation	15
Figure 1-17	Refocusing after a single exposure of the light field camera	17
Figure 1-18	Rendering different viewpoints from a single light field camera exposure ..	17
Figure 1-19	Scheme of pickup (a) and display (b) stages of IP (pinhole)	18
Figure 1-20	Scheme of pickup (a) and display (b) stages of II (microlens)	19
Figure 1-21	Example of conversion of an elemental image	19
Figure 1-22	(a) DOF of a lenslet. (b) Positions of lenslet image planes when lenslets with different focal lengths and sizes are used.	20

Figure 1-23	Elemental images with (a) classical procedure and with (b) hybrid procedure..	21
Figure 1-24	Opaque barriers are used to avoid overlapping of elemental images	22
Figure 1-25	Scheme of telecentric relay system	22
Figure 1-26	Scheme of multiple-axis telecentric relay system	23
Figure 2-1	Depth cues of human visual system	28
Figure 2-2	Binocular vision: 2D stereo image pair fused into 3D cyclopean image	28
Figure 2-3	Geometry of binocular vision	29
Figure 2-4	Oculomotor depth cues: accommodation and convergence	30
Figure 2-5	Numerical aperture of an optical system	31
Figure 2-6	Rayleigh and Sparrow criteria for two overlapping diffraction patterns	32
Figure 2-7	Depth of focus and depth of field	34
Figure 2-8	An unaided view of an arrow object	35
Figure 2-9	An aided view through a magnifying glass	36
Figure 2-10	A rudimentary compound microscope	37
Figure 2-11	Basic optical transmission microscope and its elements	38
Figure 2-12	Dodging and burning effect	40
Figure 2-13	Example of high dynamic range image by tone mapping	41
Figure 3-1	Scheme of extended depth of field	43
Figure 3-2	Spatial HDDR system with 2 DOF	44
Figure 3-3	Temporal HDDR system with 2 DOF	44
Figure 3-4	HDDR system with 3 DOF	45
Figure 3-5	Experiment setup	46
Figure 3-6	Effective lens design	47
Figure 3-7	Overall imaging system with lens array	48
Figure 3-8	Flow chart of algorithm	49
Figure 3-9	Example of results from moves in graph cut algorithm.(a) Original pixel	

labeling (b) after an α -expansion move	52
Figure 3-10 Deviation correction	53
Figure 3-11 Representative point of a deviation-corrected elemental image.....	54
Figure 3-12 Scheme of threshold determination	55
Figure 3-13 Segmentation voids	56
Figure 3-14 Three situations while reconstruction (a) ideal image (b) voided image	56
Figure 3-15 Examples of median filtering of three situations of reconstruction (a) object reconstruction (b) background reconstruction (c) error reconstruction.....	57
Figure 4-1 Elemental images under F/22 with different exposure time (a)(b)(c) three perspectives with adequate EV, (d)(e)(f) one tenth of adequate EV, (g)(h)(i) adjusted images of (d)(e)(f) respectively	59
Figure 4-2 Rendered depth map from under-exposed elemental images	60
Figure 4-3 Blurred image and its depth map (a)(d) ground truth (b)(e) 21-pixel variance (c)(f) 41-pixel variance.....	60
Figure 4-4 Error rate versus the variance of disk	61
Figure 4-5 Three out-of-focus elemental images (a) left perspective (b) central perspective (c) right perspective.....	61
Figure 4-6 Rendered depth map from blurred elemental images	61
Figure 4-7 Six elemental images of three perspectives and two focal positions and captured under F/2.8 (a)(c)(d) focus at foreground (b)(e)(f) focus at background	63
Figure 4-8 Two rendered depth maps (a) focus at foreground (b) focus at background....	63
Figure 4-9 Details of objects in rendered depth maps (a)(c) in focus (b)(d) out of focus..	64
Figure 4-10 Experiment images during fusion process (a)(b) finding representative focal point (c) two depth maps fusion after thresholding.....	65
Figure 4-11 Depth maps rendered of (a) HDDR system (b) large f-number (f/22)	65
Figure 4-12 Elemental images of three focal positions and captured under F/2.8 (a)(b)(c)	

focus at first object (d)(e)(f) focus at middle object (g)(h)(i) focus at the last object from left, central and right perspective respectively	66
Figure 4-13 Three rendered depth maps (a) focus at first object (b) focus at middle object (c) focus at the last object	67
Figure 4-14 Details of objects in rendered depth maps (a)(d)(g) focus at the first object (b)(e)(h) focus at the middle object (c)(f)(i) focus at the last object	68
Figure 4-15 Experiment images of finding three representative focal points	69
Figure 4-16 Experiment results of fusion and its details	70
Figure 4-17 Depth maps rendered of (a) HDDR system (b) large f-number (f/22)	70
Figure 4-18 Comparison of the first object in color image and depth map of (a)(c) HDDR system (b)(d) large f-number (f/22)	70
Figure 4-19 Variance of different F/# versus depth	71
Figure 4-20 Working range of different focal designs	72
Figure 4-21 Exposure time of different focal designs	72
Figure 4-22 Six elemental images of three perspectives and two focal positions and captured under F/2.0 (a)(c)(d) focus at foreground (b)(e)(f) focus at background	74
Figure 4-23 Two rendered depth map and the result after fusion (a) focus at foreground (b) focus at background (c) HDDR depth map	75
Figure 4-24 Six elemental images of three perspectives and two focal positions and captured under F/2.0 (a)(c)(d) focus at foreground (b)(e)(f) focus at background	76
Figure 4-25 Two rendered depth map and the result after fusion (a) focus at foreground (b) focus at background (c) HDDR depth map	77
Figure 4-26 Matching range and blind range in near field	79
Figure 4-27 Occlusion geometry	79
Figure 5-1 Mechanism of GRIN lens	83
Figure 5-2 Scheme of the 3D model of a tumor growing	84

Figure 5-3 Horizontal scheme of fine depth resolution design 85

Figure 5-4 Distribution of lens array with different rendering depth ranges 85



List of Tables

Table 1-1	Summary of 3D capturing techniques.....	25
Table 3-1	Experimental Parameters	46
Table 4-1	Comparison of two HDDR systems.....	78
Table 5-1	Comparison table of our HDDR system with the prior arts.....	82



Chapter 1

Introduction

1.1 Preface

Have you ever noticed that your life had been invaded since camera was invented? Without a doubt, photography subtly helps us memorize the good old days like Figure 1-1.



Figure 1-1 Good old days

Formerly people were definitely the characters in the pictures because the negatives were valuable at that time. However thanks to Steven Sasson, the advent of digital camera in company with flat panel display approves people to capture the beauty more freely [1]. We can “paint” the photos like drawing on a board and don’t be afraid of making any mistakes. As a result, anything meaningless appearing on the picture seems reasonable. Moreover, the style of dilettante is catching on quickly due to the wide spread of digital camera as well.

Figure 1-2 shows some examples of the style of dilettante, in which the photo gives an artistic conception to the significant sentences.



Figure 1-2 The style of dilettante

Digital camera provides many potentialities while taking a picture, but does the scene in a picture frame represents your entire world? How about the additional freedom of capturing, depth?



Figure 1-3 The evolution of displays

Actually, display devices tactfully guide the evolution of camera. Figure 1-3 seems to reveal the next generation of digital camera as if smart phones almost take the place of conventional cell phones due to the improvement of micro processors as well as the touch function. Accordingly what would be displayed on the displays that are capable of projecting three-dimensional images? Is it possible to record your own 3D memory by yourself?

As far as I am concerned, 3D image capturing, display, or even interaction technology is devoted to make the audiences or customers feel like they are standing right in the scene itself. Because of the spirit of human technology, autostereoscopic 3D display is a leading target that many scientists and engineers hope to achieve. Likewise, air touch of bare finger is another promising concept, not to speak of the essential 3D content. From above, our prospect of virtual reality is shown as Figure 1-4. Hence in this thesis, a practical optical system for 3D image capturing is proposed. Furthermore, considerable prior arts and fundamental theory give an insight and the background of 3D technology and related optics. Last but not least, the feasibility of the proposed system is discussed on the basis of the functionality of optical lens.

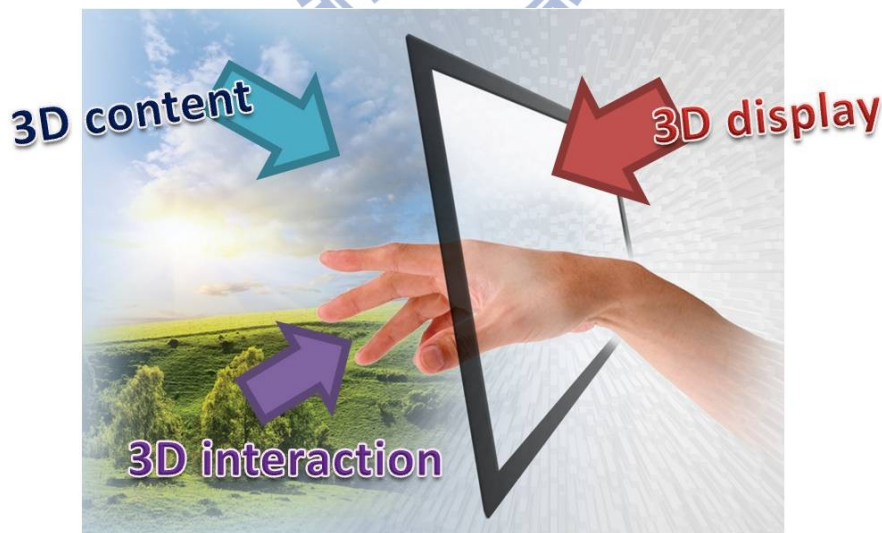


Figure 1-4 Prospect of virtual reality

1.2 Prior Researches

It goes without saying that prior arts are the cornerstone of a research. Isaac Newton has said that *“If I have seen further it is by standing on the shoulder of a giant.”* As a result, it is inevitable to introduce some techniques for acquiring depth information. Researchers are trying to analyze how people view the 3D world and to imitate the human visual system to reconstruct the depth from a single 2D image or plenty of 2D images. Figure 1-5 illustrates the categorization of those rendering systems with three classes of computer-based, multi-camera, and single-camera types. All of them are outstanding methods but not perfect somewhere.

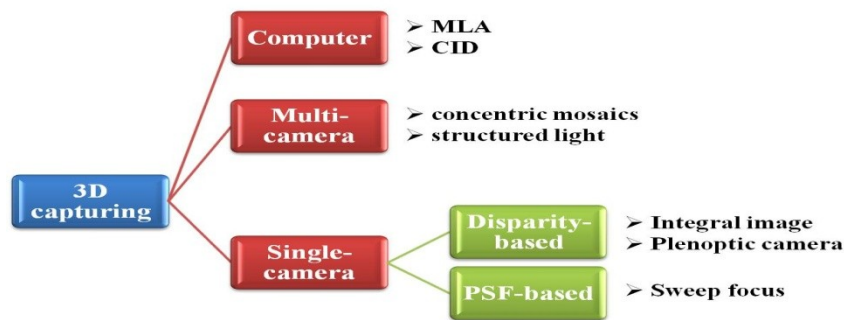


Figure 1-5 Categorization of 3D capturing techniques

1.2.1 2D-3D Conversion

Intuitively, it is straightforward to obtain the 3D information if we could implement some image processing directly onto existing abundant 2D materials. 2D-3D conversion is exactly the computer-based technique, which estimates the depth information from a single or a sequence of 2D images. There are two categories making use of the correspondence between subsequent frames: Depth from Motion (DFM) and Structure from Motion (SFM). Modified

Time Difference (MTD) method based on Pulfrich Effect [2] is one kind of 2D-3D conversion belonging to DFM. MTD selects the stereo pairs by detecting the horizontal motion to develop 3D perception from input sequential images, as shown in Figure 1-6.

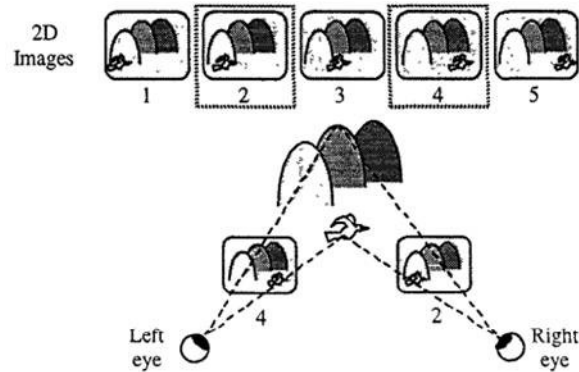


Figure 1-6 Principle of MTD

In the case of SFM, similar views of the same scene avail determining the depth. But this approach needs all objects remaining stationary. In short, these techniques relying on the correspondences between frames are unreliable as the scenes involves low textured and fast moving object. Moreover, they fail in recovery of depth in the absence of any motion however. As a consequence, other more pragmatic methods are proposed to solve the problem of 2D-3D conversion. For instance, Machine Learning Algorithm (MLA) is used for generation of depth map. MLA method usually consists of two stages: training and classification. During the training stage, the color of individual pixels is input associated with known depth. MLA then adjusts its internal configuration to learn the relation. As long as we apply this relation to an untrained sample, an output depth value will be determined. Aside from MLAs, Computer Image Depth (CID) method is another manner capable of transforming all kinds of 2D images into 3D images according to contrast, sharpness and chrominance. Generally, higher contrast and sharpness stand for near-positioned objects. Notwithstanding CID is especially suitable for converting from still images, CID cannot produce a stereo-occlusion yet. [3][4] Regardless of MLA or CID methods, they are advanced but semi-automatic. [5] has proposed a fully automatic method to generate depth map from a single input image. The input image is

processed by the following steps:

- a. Bayer to approximated-RGB color conversion
- b. Color-based segmentation
- c. Ruled-based regions detection to find specific areas
- d. Image classifications to discriminate between outdoor with and without geometric elements and indoor images
- e. Approximated depth map estimation

At first, the dimension of color space is reduced. Via mean shift algorithm, image is under-segmented basing on likeness of the color of each pixel. Then the following is the most significant step, region detection. Semantic areas are classified into six classes such as sky, mountain or land, by means of color-based rule. Each specific region is assigned different grey level values and become the qualitative depth map. In addition, geometric information is analyzed to generate a geometric depth map which roughly allocates the direction of extending depth. In the end, two maps, qualitative depth map and geometric depth map, are fused together into final depth map, as shown in Figure 1-7.

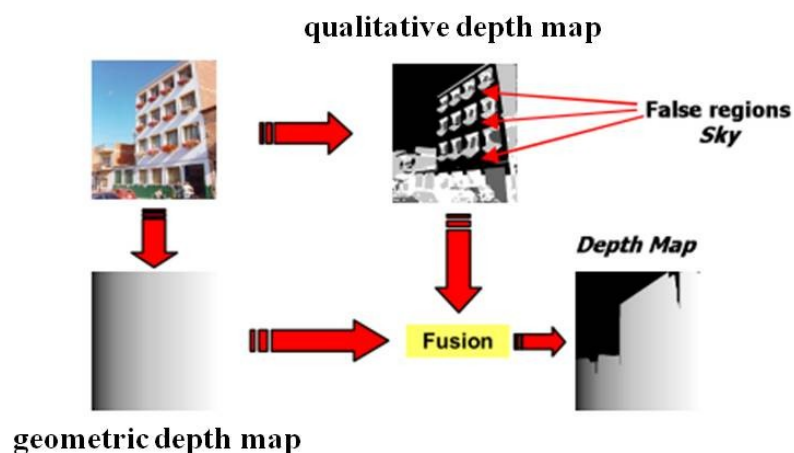


Figure 1-7 Generation of depth map by fusion

However, several experimental results have confirmed the robustness of this automatic method but the extension to further image categories is definitively needed. [5]

1.2.2 Multi-camera System

Image-based rendering techniques have been developed to accomplish free viewpoint images from a set of pre-acquired images. These techniques create photorealistic images based on plenoptic function which includes recording every position, angle, wavelength and time instant of real world. There have been diverse approaches proposed via plenoptic function of different dimension. Concentric Mosaics represent 3D plenoptic function and capture a scene by spinning an off-centered camera on a rotary table. Although concentric mosaics have much smaller file size than Lumigraph, it lacks vertical parallax. This weakness brings about occlusion issue. As illustrated in Figure 1-8, we cannot reproduce view rays like $P_{view}F$ because the camera doesn't capture the part around point F.

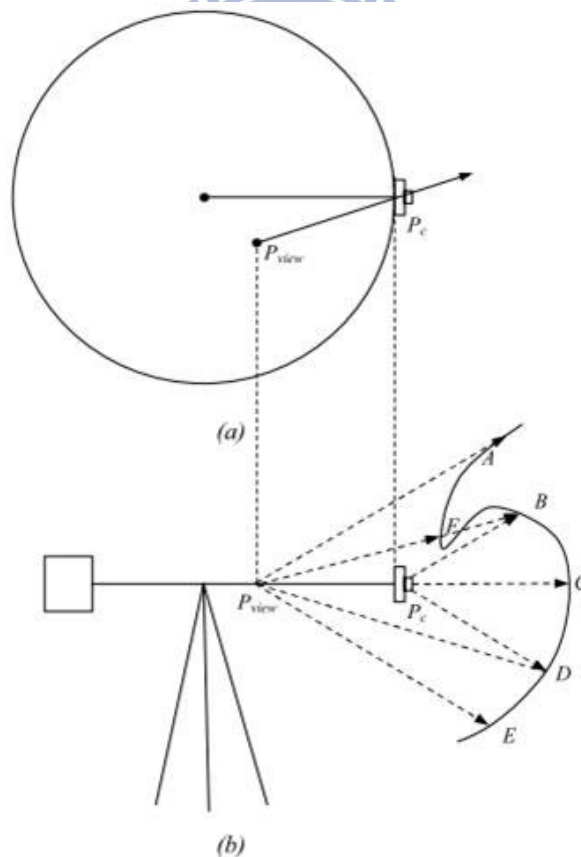


Figure 1-8 Scheme of concentric mosaics method, (a) top view (b) side view

Therefore, concentric mosaics system is extended to longitudinal aligned camera array system depicted in Figure 1-9 so as to retrieve the rays off the capture plane. Given the viewing position and direction of an observer, a novel view can be interpolated accordingly. Camera array indicates supplementary data; however, if the viewers are constrained to move on a plane, this system doesn't need to store all the captured rays. So, the file size is almost equivalent to a 3.5D plenoptic function. But even for the same situation, a 4D plenoptic function still requires 2D for position and 2D for ray direction. [6]

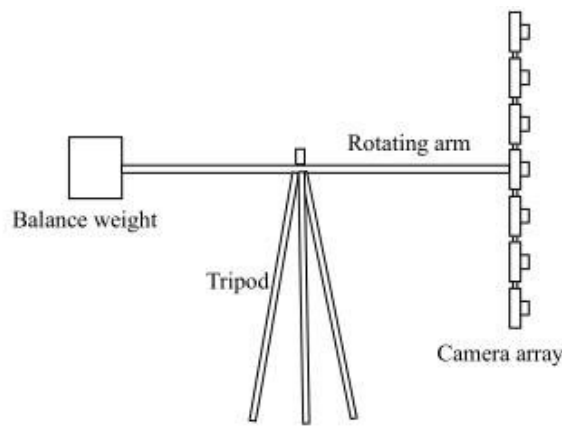


Figure 1-9 Setup of longitudinal aligned camera array system

In addition to generate a free viewpoint image, dense entire shape acquisition is also required. Shape from silhouette and Multi-view stereo are techniques that have commonly been used. However, recovery of concavity and small bumps hold silhouette method back while if scene texture is uniform, it is difficult for multi-view stereo to retrieve precisely. Hence active 3D scanning is used for practical purposes owing to its accuracy and fidelity. How 3D scanning works is to capture the scene with structured light projected. Either temporal-encoding-based projected-camera system or spatial-encoding type, they are confined if there's only one camera being used. Although temporal-encoding type is popular now, basically it takes longer time to achieve entire shape acquisition. On the other hand, complicated patterns cause a predicament for spatial-encoding system to decompose well, even though different colors are applied. Consequently, spatial-encoding system of single

camera results in not only sparse reconstruction, but also failure in reconstruction of small objects [7]. Aforementioned approach has succeeded only in static object, so [8] extends the origin system, six cameras and projectors, to dynamic moving bodies.

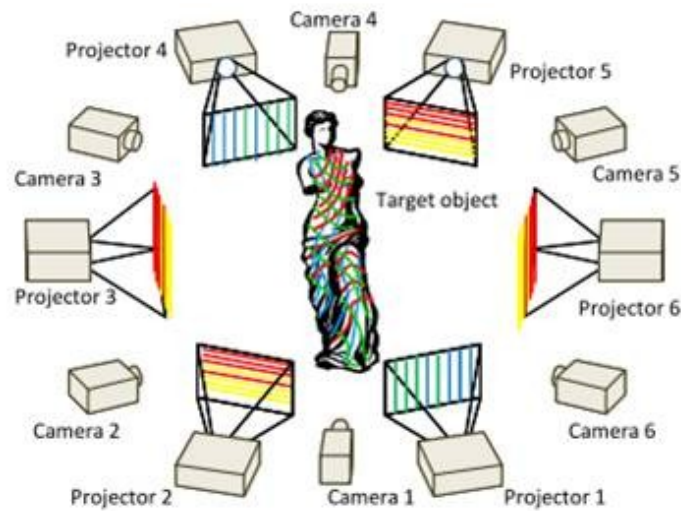


Figure 1-10 Setup to reconstruct the entire shape

Figure 1-10 illustrates the configuration that cameras and projectors are placed alternatively so that each camera can observe the parallel single direction lines projected by adjacent projectors. Observing the intersection points of the lines is a major part for reconstructing the shape. [8] Lest abstruse algorithm will confuse you, detailed algorithm is not discussed here.

In spite of the fact that multi-camera system in different configurations will undoubtedly capture abundant information of the real world, accordingly it is favorable in film industry. Nevertheless, camera array implies it is cumbersome and goes against consuming products.

1.2.3 Sweep Focus

Before taking a picture, photographers have to select particular parameter setting to enhance different ambiances in the photos. Depth of field is one major factor used to emphasize some specific objects by defocusing the foreground and background. However, with the advent of computational photography, viewers are allowed to explore the scene after the picture is captured. For example, people can interactive control the depth of field of a photograph, which is referred to refocusing. The basic representation used to facilitate refocusing is focal stack, or termed as space-time focal volume, as depicted in Figure 1-11. Instead of computing focal stack from light field, it is more intuitive as shown in Figure 1-12 to physically sweep its focal plane across a scene. Such a system is called focal sweep camera.

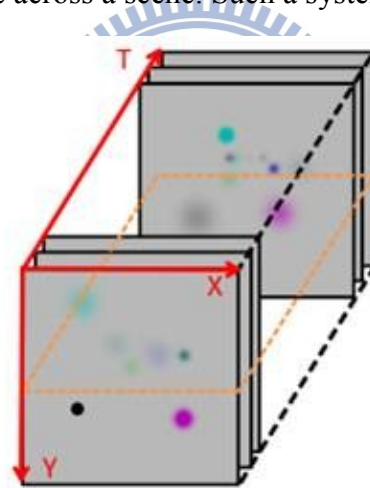


Figure 1-11 Focal stack or space-time focal volume

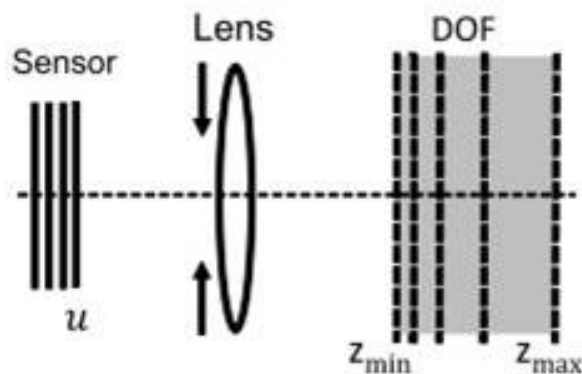


Figure 1-12 Depth of field of multiple shots

Along the optical axis, T-axis, each ball appears as a double-cone. By detecting the apex of every double-cone, all-in-focus image can be rendered. However, larger motion makes the scene too blur to analyze the focal stack. Therefore the capturing time should be minimized to reduce motion blur. With an eye to completeness and efficiency, aggregated depth of fields should cover entire range but not overlap with each other. As a result, the sampling strategy is concluded that sensor should be moved by a constant distance, twice the pixel size multiplying f-number, between each consecutive image. [9]

In order to extend the depth of focus, besides mechanical moving of sensor or specimen, the intrinsic problem, chromatic dispersion of an imaging system, however seems analogous to the same concept. Because refractive index is a function of wavelength, hence according to Snell's Law, the focal length also varies as a function of wavelength, as shown in Figure 1-13. Such a system is called Spectral Focal Sweep (SFS) camera and the amount of focal sweep depends on the reflectance spectra of objects. Fortunately, the reflectance spectra are sufficiently broad in real world. Moreover, large f-number is required without sacrificing the signal-to-noise ratio (SNR). [10]

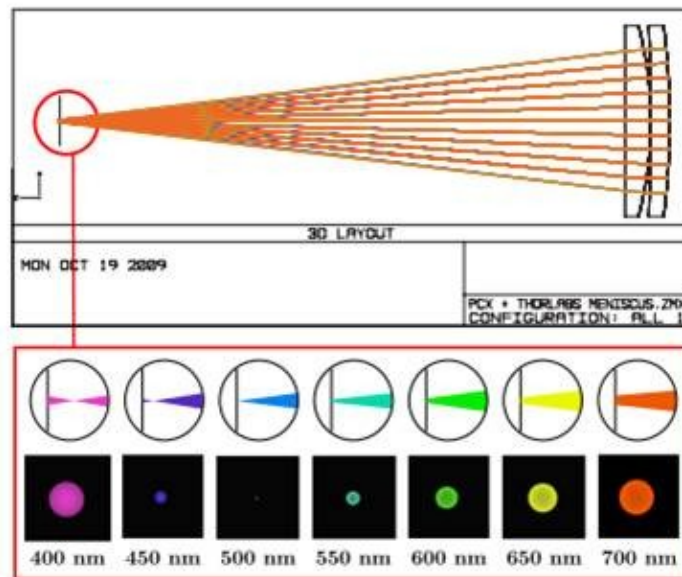


Figure 1-13 Chromatic dispersion of focal lengths

Compared to light field camera, it takes a finite duration of time for focal sweep camera to capture all the images, but it benefits views to perceive scene dynamics along with depth while refocusing. On the other hand, focal sweep camera preserves sensor spatial resolution because light field camera contains the dimensionality gap between captured and required information. As for other related works, depth from focus or defocus involves ambiguities of focus measure at different scale. Besides, non-textured regions reveal nothing about depth based on focus measure.

Although focal sweep generally presents less sacrifice of sensor resolution among all single-camera system, mechanical moving and longer capturing time might be a disturbance for practical commercialization. Even though SFS camera is characterized with no mechanical moving, the reconstruction of colorful scene is sometimes challenged because it inexactly defocuses the luminance channel only. Furthermore, depth-dependent point spread function is required to distinguish the objects locating at different longitudinal positions, but this requirement cannot be achieved as those objects fall within the same range of depth of field. This limitation reveals the nonlinear depth resolution while refocusing.

1.2.4 Light Field Camera

As the concept of light field was developed by Michael Faraday in 1846 [11], two-dimensional photographic image were no longer simply a two-dimensional world. But in fact, Leonardo da Vinci had explored a similar idea with these words: “*Every body in the light and shade fills the surrounding air with infinite images of itself; and these, throughout space and on every side.*” [12]

Light field standing on the geometric optics and radiometry represents the amount of light faring in every direction through every point in space. In computer vision, light field is described as plenoptic function, from word root for “complete” and “view”, which expresses the image of a scene from any possible viewing position at any viewing angle at any point in time. At the very start, plenoptic function is five-dimensional because rays are parameterized by three coordinates and two angles. Undoubtedly additional variables, such as wavelength or polarization, yield higher dimensionality. However, so long as it is excluded that both light rays come in and hit the object and light rays emanate from the object on the opposite side, plenoptic function is reduced as a four dimension function and commonly parameterized by two 2D planes, sometimes called light slab. Note that 4D light field is not equivalent to capturing two 2D planes of information.

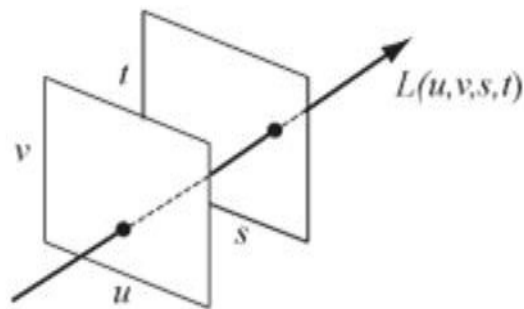


Figure 1-14 Parameterization of light slab

Instead, light slab specifies point pairs to indicate each unique ray, as depicted in Figure 1-14. Actually this light slab is carried out through a supplementary lens array. Figure 1-15

illustrates a conceptual scheme of light field camera in which main lens and microlens array are analogous to u-v plane and s-t plane respectively.

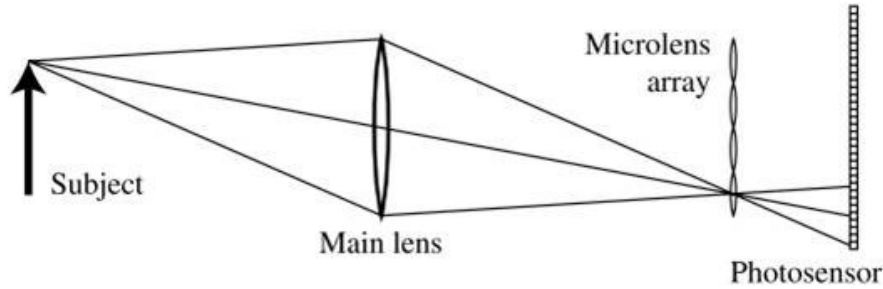


Figure 1-15 Conceptual scheme of light field camera

The main lens may be translated along its optical axis to focus on the subject of interest at desired depth. Each object point is brought to a single convergence point on the focal plane of microlens array. Subsequently, every microlens forms a tiny sharp image of lens aperture, called subimage. The subimages capture the structure light, visual pyramid, and reveal the depth of objects.

Besides, in order to maximize the directional resolution, the main lens is placed at the optical infinity of microlens and the distance between photo sensor plane and microlens array is set as the focal length of microlens. On the other hand, the directional resolution relies on the size of subimage as well, so the subimage becomes largest without overlapping when the f-numbers of main lens and microlens are equal. But the f-number of main lens is of interest in image side and defined as the ratio of the separation between the principle plane of the main lens and the microlens plane.

Regarding image synthesis, only four parameters are considered for simplicity: aperture size and location, and the depths of the microlens and sensor planes. Begun with cosine fourth law, the irradiance image value that would have appeared on the synthetic film plane is given by

$$E(s', t') = \frac{1}{D^2} \iint L'(u', v', s', t') A(u', v') \cos^4 \theta \, du \, dv \quad (1)$$

where $L(L')$ and A are the light field and an aperture function respectively, and θ is the angle

function centered at (u_o, v_o)). And because it doesn't matter where the focal plane is, α is set equal to 1 and the synthetic photograph equation is converted as

$$\bar{E}(s', t') = L\left(s' + \frac{u_o - s'}{\beta}, t + \frac{u_o - t'}{\beta}, s', t'\right) \quad (5)$$

Obviously pinhole rendering is considerably faster than refocusing since we do not have to perform the double integral over the main lens. In sum, the overall concept of image synthesis is to re-sort the rays of light to where they would have terminated. Figure 1-17 and Figure 1-18 demonstrate the refocusing and free viewpoint rendering respectively. [13]

Light field camera has a great deal of benefits in many areas. In photography, lens aberration can be diminished by re-sorting the distorted rays. Moreover, light field approves extending depth of field while using a large aperture, and benefits sports photography and security surveillance, for example. The capabilities of extending depth of field and changing viewpoints also play a leading role in medical and scientific microscopy. However, sensor resolution dominates the performance of light field because directional resolution is limited by the pixel size of sensor and diffraction effect. On the other hand, 4D light field records redundant information, which causes the image resolution decrease dramatically.



Figure 1-17 Refocusing after a single exposure of the light field camera

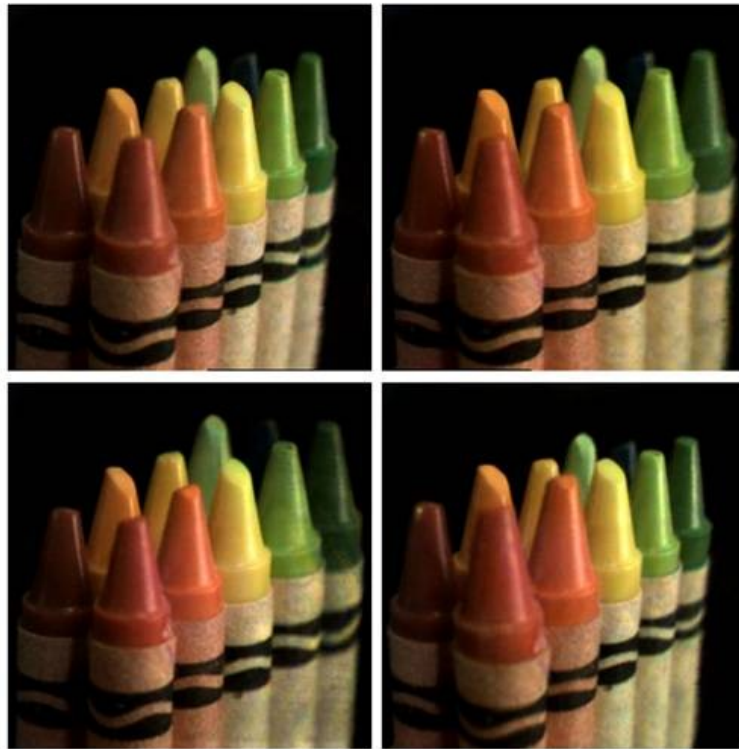


Figure 1-18 Rendering different viewpoints from a single light field camera exposure

1.2.5 Integral Image

Recording and displaying of 3D scenes has bought about many attentions of scientists. In the nineteenth century Sir Wheatstone first built the stereoscope based on the binocular disparity. Moreover, Integral Image (II) or Integral Photography (IP) proposed by Lippmann at the beginning of twentieth century allows the reconstruction of 3D images relying on the reversibility of light rays [14].

The principle of integral image includes two stages: pickup and display. In Lippmann's scheme, the configuration is composed of a pinhole array and a sensing plane. Figure 1-19 (a) illustrates that pinhole array specifies the directions of light rays in pickup stage. If the sensing plane is replaced with a photographic film in display stage, the recorded information can be delivered reversely and rebuilds the original 3D image as shown in Figure 1-19 (b).

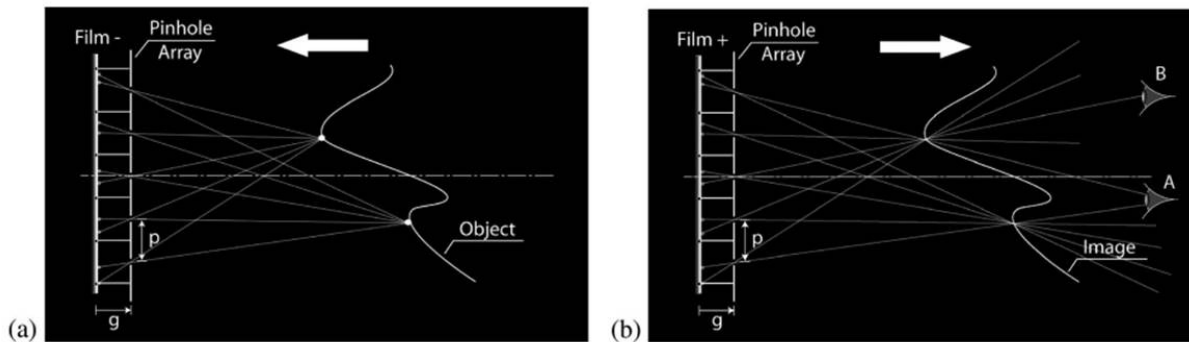


Figure 1-19 Scheme of pickup (a) and display (b) stages of IP (pinhole)

However, it is obvious that viewer observes the wrong side of the 3D scene, i.e. depth-reversed. Pseudoscopic 3D image is generated without rearranging the recorded elemental images. Besides, the major disadvantage of Lippmann's scheme is poor light efficiency of both two stages. To improve the light efficiency, increasing the aperture of pinholes will cause the deterioration of lateral and depth resolution. Hence, Lippmann's scheme is ameliorated by substituting with a microlens array and an electronic matrix sensors like a CCD or a CMOS as shown in Figure 1-20.

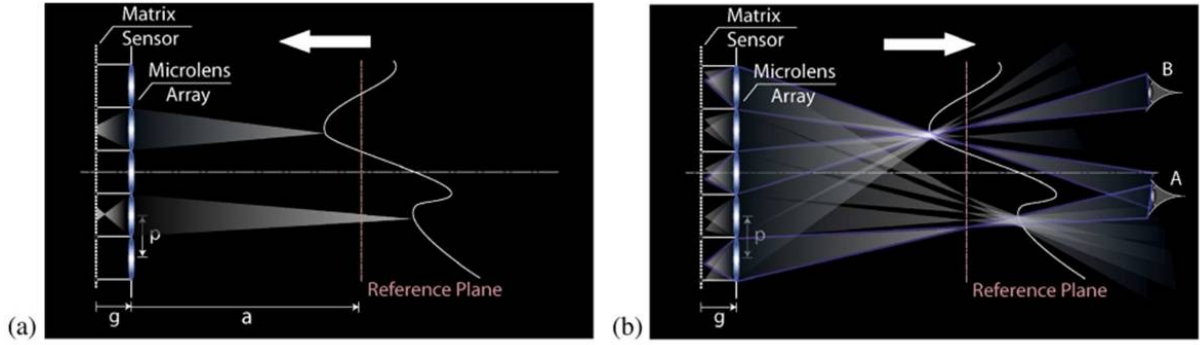


Figure 1-20 Scheme of pickup (a) and display (b) stages of II (microlens)

Even though the spatial resolution of electronic matrix sensors might be comparable to classical photographic films, image capacity of microlens leads to only a reference plane in focus. Furthermore, pseudoscopic image is also formed because the direction of pickup and viewing is opposite. To solve this issue, so-called orthoscopic-pseudoscopic image-conversion optics using specially designed prisms is proposed. However, Figure 1-21 depicts another preferred method that each elemental image is shifted centrosymmetrically about the center of the elemental image [15].

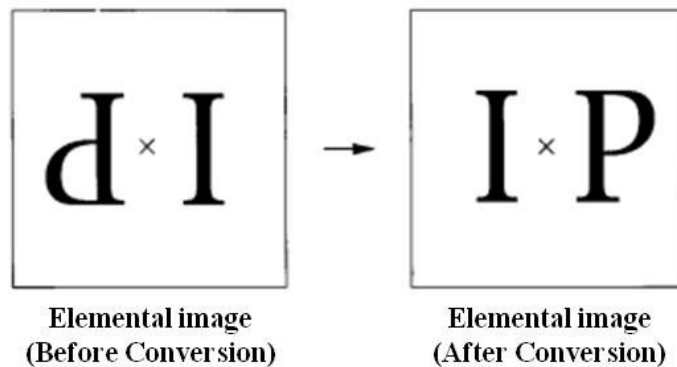


Figure 1-21 Example of conversion of an elemental image

If we believe the spatial resolution of electronic matrix sensors will reach that of photographic film, there are three main challenges of integral image in pickup stage: 1) limited depth of field, 2) overlapping of elemental images, and 3) viewing angle.

According to Fraunhofer diffraction, the minimum spot size s is given by

$$s = 2\lambda L_i / w_i \quad (6)$$

where λ is the wavelength of illumination, L_i is the distance between lenslet array and the lenslet image plane and w_i is the side length of each lenslet. Then the resolution R is the inverse of spot size.

$$R = 1/s = w_i/2\lambda L_i \quad (7)$$

On the other hand, based on Rayleigh criterion, the maximum depth limit D is

$$D = 4\lambda L_i^2/w_i^2 \quad (8)$$

Hence, regardless of lenslet size and focal length, the product of depth and resolution square (PDRS) is a constant given by

$$DR^2 = 1/\lambda \quad (9)$$

This relation indicates the trade-off between depth of field and spatial resolution although the constant is deviated with different diffraction limited cases [16][17]. As depicted in Figure 1-22, one solution for increasing the image depth without increasing the spot size is to utilize microlens array with different focal lengths and aperture size according to following equations:

$$L_{i+1} = L_i + D = L_i + 4\lambda L_i^2/d_i^2 \quad (10)$$

$$d_{i+1} = d_i + 2s = d_i + 4\lambda L_i/d_i \quad (11)$$

$$f_{i+1} = gL_{i+1}/(g + L_{i+1}) \quad (12)$$

where the focal length of lenslets and the gap distance between the sensor/display plane and the microlens array are denoted by f and g respectively.

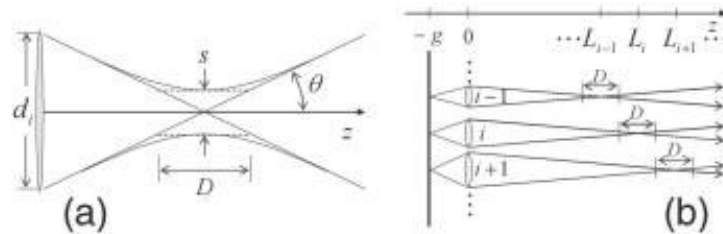


Figure 1-22 (a) DOF of a lenslet. (b) Positions of lenslet image planes when lenslets with different focal lengths and sizes are used.

The resolution of a 3D image however is governed not only by the spot size but by the lens pitch. By Nyquist ray sampling rate, image depth is improved at the cost of decreased resolution [18]. To overcome any severe deterioration of lateral resolution, a circular obscuration at the central region of lenslets is inserted to increase the depth of field because the elemental PSFs spread very slowly for out-of-focus points. Therefore, deconvolution such as Wiener deconvolution algorithm can be applied to enhance the image quality as shown in Figure 1-23. However, high pass filter implies low light efficiency.

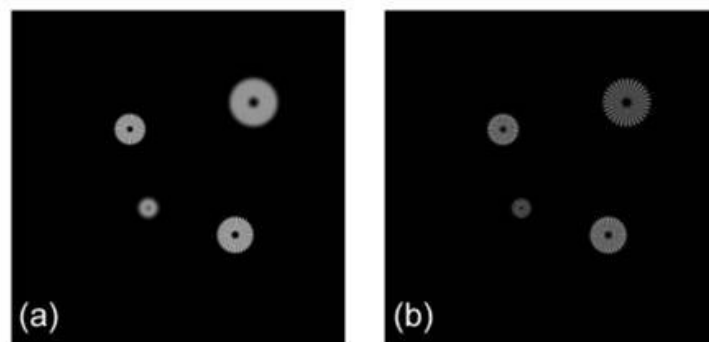


Figure 1-23 Elemental images with (a) classical procedure and with (b) hybrid procedure

Overlapping of elemental images stems from the oblique imaging. A typical approach is to insert opaque barriers as shown in Figure 1-24. As for purely optical method, Telecentric Relay System (TRES) is proposed by controlling the micro entrance pupils (micro-EPs), the images of the aperture stop through the microlenses. Figure 1-25 illustrates that the regions of image formation is distinguished beforehand. Both of the two ideas derive from same concept but operating at either object side or image side. Hence the image will be reconstructed more clearly but lose some light illumination.

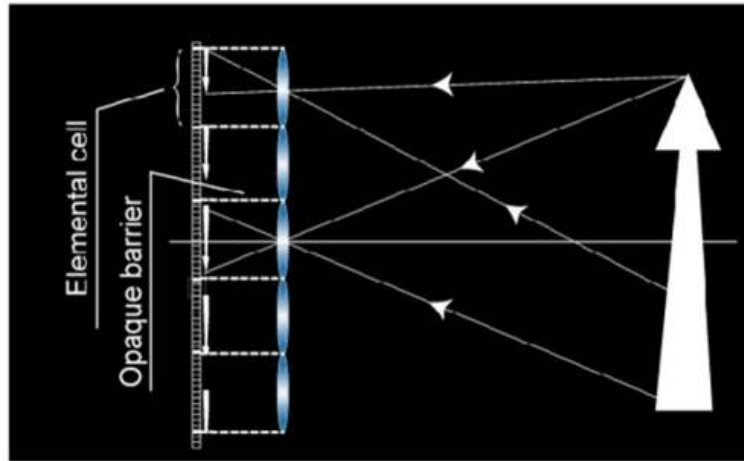


Figure 1-24 Opaque barriers are used to avoid overlapping of elemental images

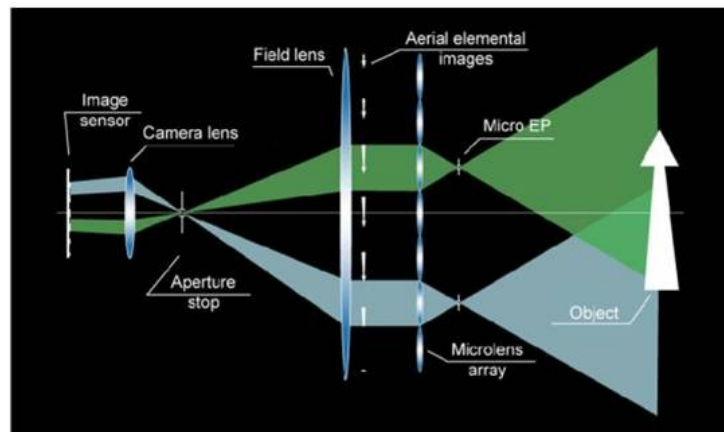


Figure 1-25 Scheme of telecentric relay system

While oblique imaging is blocked, the viewing angle is simultaneously confined because of the deficiency of recorded image information. As a consequence, it is straightforward to join parallel micro-EPs of each lenslet for compensating the angular data. Multiple-Axis Telecentric Relay System (MATRES) modifies TRES with a camera array as depicted in Figure 1-26. The central camera captures the same collection of elemental images, but the left and right cameras acquire additional elemental images with complementary perspectives. In other words, each micro-EP has three cones to form the image and yields a threefold increase of the field of view. However, supplementary sensing elements not only induce a cumbersome system but also complicate the process. [19]

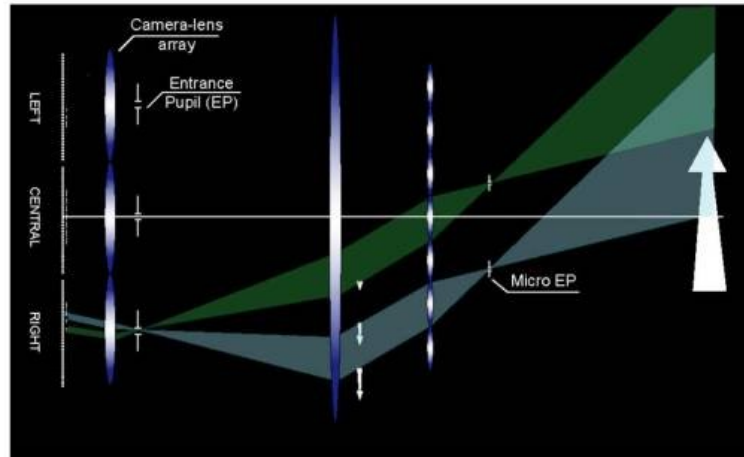


Figure 1-26 Scheme of multiple-axis telecentric relay system

Integral image is a promising technology by providing numerous advantages: 1) it produces autostereoscopic 3D images with full parallax and continuous viewing points, 2) it avoids convergence-accommodation conflict, 3) it can be operated with incoherent light, and 4) it is compact and easy to implement. Nevertheless, the tremendous sacrifice of sensor resolution is a vital issue not only for image quality but for accuracy of depth computation.



1.3 Motivation and Objective

Three-dimensional technology is blooming in our daily life from 3D TVs to 3D movies. People pursue a virtual reality to experience an environment that you have even never been to. But to reach the new era of displaying or entertainment, abundant 3D contents are undoubtedly in great demand.

There are many approaches to fulfill the transformation of conventional 2D images into 3D form used in 3D displays. However, 2D-3D conversion techniques narrow the application of 3D movies due to the fact that some image categories lack reliable accuracy. Moreover, since people are used to delicate 2D cinema, the film quality should be the first priority. As a result, multi-camera system is usually utilized in film industry on account of plentiful image information for reconstruction of 3D images. Notwithstanding multi-camera system ideally will provide the best characteristics of 3D movies, the cumbersome scheme leads to not only complex calibration but also the obstruction of commercialization. With regard to consuming products or even portable devices, single-camera system is worth being taken into consideration. Sweep focus, light field camera, integral image are three favorable candidates of this system. Among them, sweep focus preserves highest sensor resolution of elemental images, but it takes time to scan along the depth to construct the full focal stack. Mechanical moving is also another torment for consumers. In addition, sweep focus relies on depth-dependent point spread function, so the depth resolution is limited for those objects in the same depth of field. As for light field camera and integral image, they make use of 4D light field to capture a 3D scene, which implies the redundancy of recorded information. The sacrifice of sensor resolution dramatically deteriorates the quality of their displaying image. Furthermore, the feasibility is restricted owing to the shallow depth of field as the object distance becomes shorter, so the microlens array of same focal lengths cannot provide the two kinds of system sufficient image data anymore when capturing in near field.

Techniques		Pros	Cons
2D-3D conversion		<ul style="list-style-type: none"> ◆ Directly transform ◆ No additional equipment 	<ul style="list-style-type: none"> ◆ Limited image categories ◆ Relative depth values
Multi-camera system		<ul style="list-style-type: none"> ◆ Abundant image information ◆ Fine depth resolution 	<ul style="list-style-type: none"> ◆ Cumbersome configuration ◆ Calibration
Single-camera system	Integral image	<ul style="list-style-type: none"> ◆ 3D capturing and displaying 	<ul style="list-style-type: none"> ◆ Need all-in-focus elemental images ◆ Record redundancy information
	Plenoptic camera	<ul style="list-style-type: none"> ◆ No mechanical movement ◆ Single shot ◆ Work in far field 	<ul style="list-style-type: none"> ◆ Need all-in-focus elemental images ◆ Record redundancy information
	Sweep focus	<ul style="list-style-type: none"> ◆ No resolution sacrifice ◆ Work in near field 	<ul style="list-style-type: none"> ◆ Mechanical movement ◆ Long capturing time ◆ Sparse depth resolution in far field

Table 1-1 Summary of 3D capturing techniques

According to the pros and cons listed in Table 1-1, we desire to design a system that is capable of estimation depth not only within the same depth of field but also in a large range of depth. At first, we implement stereo matching method to generate the depth, so elemental images with disparity and in focus are demanded. If the elemental images are getting blurred,

feature points will be lost, which degrades the performance of stereo matching algorithm. However, photographer might think of increasing the f-number of lens to extend the depth of field, but the captured intensity cannot be maintained while using the same exposure time as well as reduced aperture at the same time. As a consequence, the image information such as contrast will be abandoned if the image is under-exposure. Besides, dimming environment or the scene with fast moving objects will both worsen the image quality while increasing the f-number. Accordingly, we utilize different focal arrangement to conceptually extend the depth of field instead of adjusting the f-number. In other words, we stack the depth of field by combining several depth maps. Inheriting the concept of High Dynamic Range (HDR) image, which will has more detail descriptions in section 2.3, we denominate our fused depth map as High Dynamic Depth Range (HDDR) depth map due to the elongation of depth rendering.

1.4 Organization of the thesis

This thesis attempts to cover whole adequate knowledge that helps you understand the core value of our proposed system. Therefore, you are recommended to begin with chapter 2 if you have no ideas about the formation of 3D images or fundamental optics. Based on the concepts elaborated in chapter 2, prior arts of 3D image capturing are introduced in chapter 1. It should be noted that image processing used in prior arts are not carefully expounded lest those contents might confuse readers to catch on purpose of our design. Moreover, the preface directs you to a specific world of 3D technology, which facilitates people to grasp the motivation and objective in chapter 1. Following chapter 1, the optical system and algorithm are revealed according to the objective in chapter 3. Subsequently, chapter 4 comprises the experimental result of captured elemental images and computed process as well as the final depth map. Furthermore the discussion is included in chapter 4. As for chapter 5, conclusion is yielded and it is discussed in future work that how liquid crystal lens avails the 3D image capturing system.

Chapter 2

Theory

The conception of 3D capturing, as the name implied, consists of 3D image formation and optical capturing. In this chapter, how people perceive depth will be introduced psychologically and physiologically. In addition, optical background of imaging is included in the part of microscopy which is an extreme case of imaging and help readers to catch on every possible related optical term. In the end of this chapter, we'll expound the idea of High Dynamic Range (HDR) image because it would link the idea of our depth rendering system.

2.1 Principle of 3D images

From the viewpoint of evolution, three-dimensional perception is an essential specialization for predators. Predators, such as lions or leopards, have to draft their hunting plans in which distance estimation based on visual sense is one major factor. The more precisely predators estimate the distance, the more likely they can catch the prey successfully. [20] Human being is one kind of predators, and how people perceive the third-dimensional information is of interest. At present, because images reaching our eyes only display two-dimensional spatial relationship, researchers have summarized many depth cues that are thought to be used in the brain for human visual system (HSV) [21][22], as shown in Figure 2-1.

Due to the fact that human eyes are horizontally separated, two images fallen onto each eye are hence with slightly difference. This difference, binocular disparity, is referred to compute the depth in the brain and fuse the two images into a 3D scene. This mechanism is named stereopsis [23][24] which is one principal way for men to reconstruct the depth, so our brains can combine stereo image pair together into the cyclopean image, if displays can deliver them to two eyes respectively, as shown in Figure 2-2 [25].

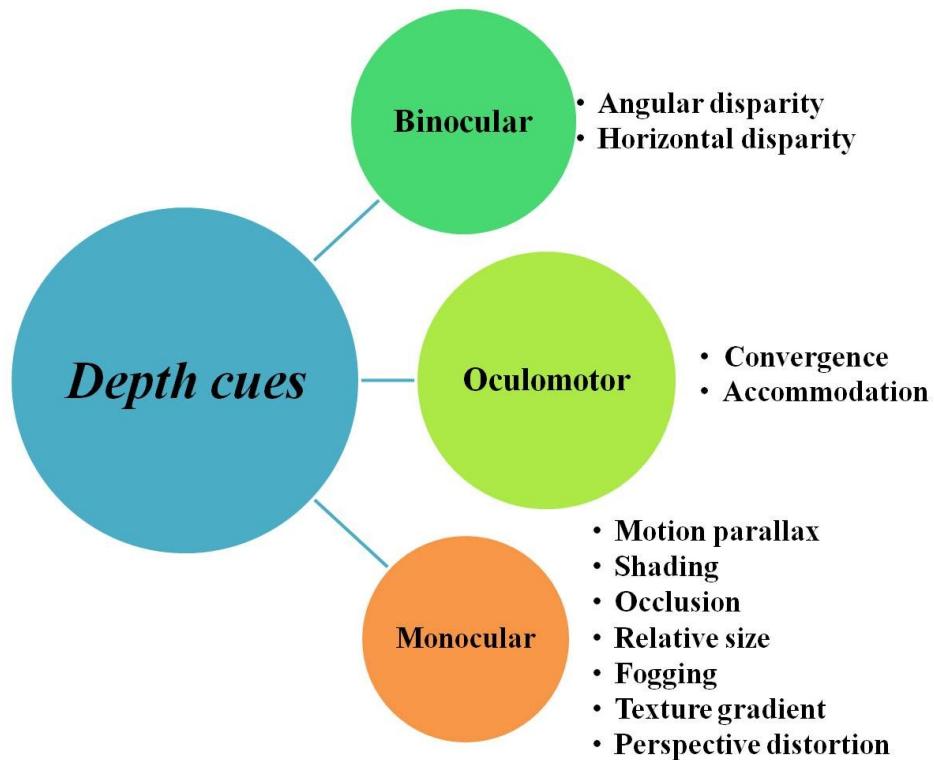


Figure 2-1 Depth cues of human visual system

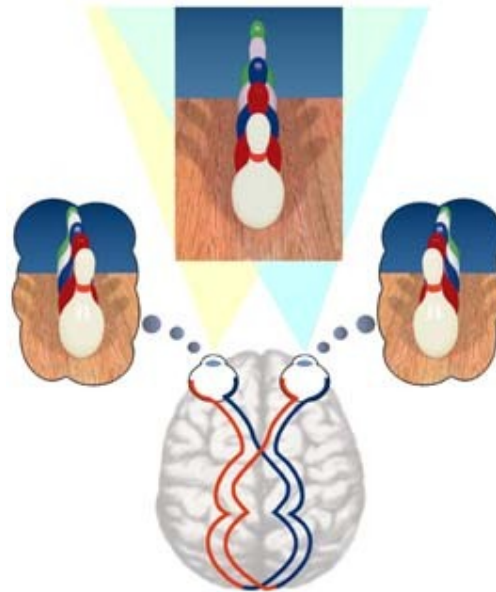


Figure 2-2 Binocular vision: 2D stereo image pair fused into 3D cyclopean image

However, quantitative measure of binocular disparity is quite important for designing 3D displays and 3D image capturing. Nowadays, 3D displays including parallax type [26] or lenticular lens [27] type, all need to reorganize the arrangement of pixels to produce stereo image pairs, and how to reorganize pixels mainly stands on the geometric relation of disparity

and depth. Likewise, concerning the reverse process, if there is a system able to shot at least two images with disparity, depth information can be computed correspondingly. This reverse process is termed 3D image capturing. Figure 2-3 illustrates the geometry when two cameras fixate at distance u' , i.e. the elemental images have no disparity.

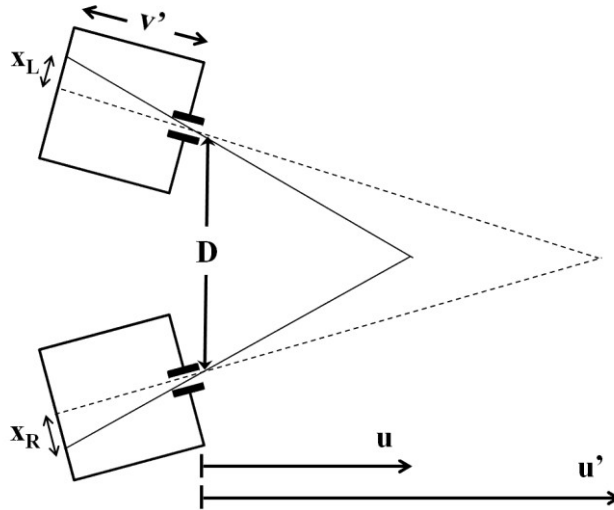


Figure 2-3 Geometry of binocular vision

Via simple geometric triangulation principle, if convergence angle is small (e.g., $D \ll u$), we can approximate an equation of disparity d at depth u where v' is the perfect imaging distance, D and F are deviation of two apertures and focal length of lens respectively [28]. As a consequence, sensitivity of depth is restricted by the size of sensing elements, i.e. pixel size, should be smaller than disparity.

$$d = x_R - x_L = D \frac{uF - v'u + Fv'}{uF} \quad (13)$$

According to above equation, we can determine the working range and setup of the capturing system; on the other hand, depth u can be rendered by the disparity of two or a sequence of elemental images when applying the same equation after rearranging it [29].

$$u = \frac{DFv'}{dF - DF + Dv'} \quad (14)$$

Besides, depth cues can be further categorized physiologically and psychologically. Physiological cues are related to eyes' motions while psychological cues are regarded as

interpretation according to previous experience [30]. As a result, the difference between oculomotor and monocular depth cues can be comprehended from physiology and psychology. Oculomotor depth cues comprise two mechanisms: accommodation and convergence as shown in Figure 2-4 [31]. Ciliary muscle and ciliary body control the optical power of lens and bring about auto-focusing in near region (~3m) which is called accommodation [32]; convergence is a process that oculomotor nerve participates in control of eye movement which helps human focus the object(s). As for monocular depth cues, Figure 2-1 only enumerates some of them but the answer to whether we could feel 3D perception by single eye can be deduced from those cues. Monocular depth cues only indicate the pretended spatial relationship of objects in the scenes by what we have built in our vision memory. In other words, even if we view a 2D picture, we know which objects are in front (3D information) but these objects actually all lie on a 2D plane. Although the details are not discussed, it is not difficult to grasp the ideas through individually visual experience [33][34].

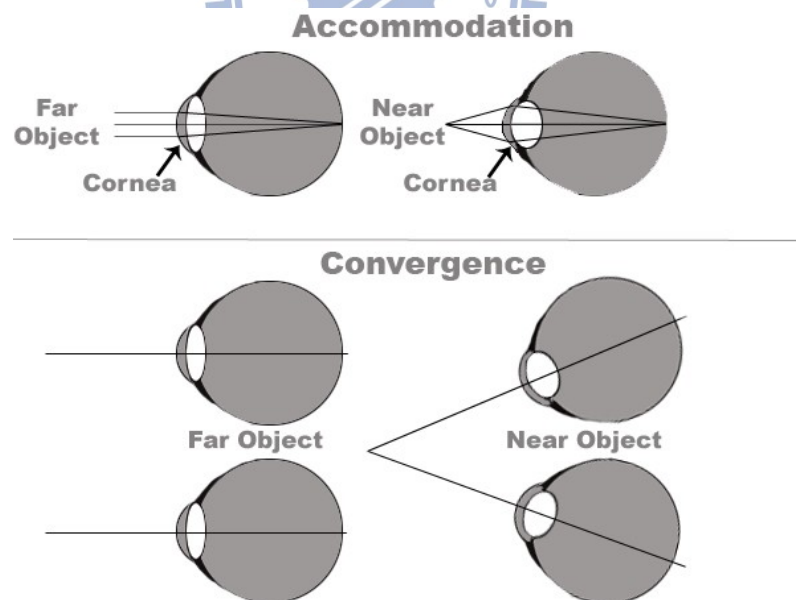


Figure 2-4 Oculomotor depth cues: accommodation and convergence

To make a long story short, the product of 3D images is however composed of more or less independent cues; to put it differently, it is not ascribed to one particular system but a step of reasoning.

2.2 Microscopy

There are plenty of things that cannot be seen with the unaided eyes, but these tiny objects do change our living somehow. Epidemical viruses could lead to coughing or a running nose, for example. Therefore, scientists and engineers have invented several kinds of instruments to investigate this mystic world. Optical, electron, and scanning probe microscopes are three prominent branches in this technical field, called microscopy [35]. Among the three branches, optical microscope is the basis of our system proposed in this thesis. Before we catch on the mechanism of it, we had better review some optical terms.

2.2.1 Optical Terminology

In general optics, **numerical aperture (NA)** is defined by the index of refraction n' (of the medium in which the image lies) times the sine of the half angle of the cone of illumination U' [36], as shown in Figure 2-5.

$$\text{Numerical Aperture (NA)} = n' \sin (U') \quad (15)$$

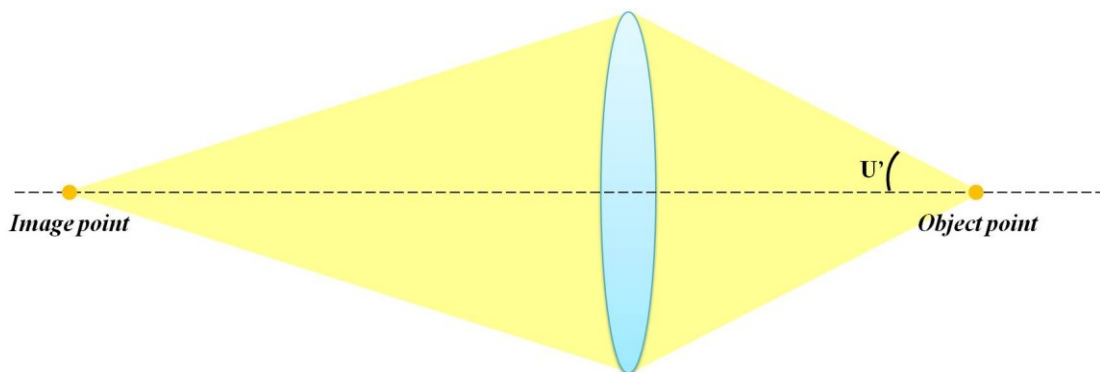


Figure 2-5 Numerical aperture of an optical system

NA is a dimensionless number used in microscopy to disclose how much light will be accepted by the system according to a particular object and it varies as the point moves. Besides, in the light of not only **Rayleigh criterion** invented by Lord Rayleigh [36] but also

Sparrow criterion invented by Carroll Mason Sparrow [37], the minimum resolvable separation **R** is inverse proportional to **NA** where λ is the wavelength of light source as the under equations of two criteria respectively.

$$\text{Resolution (R)} = \frac{0.61\lambda}{\text{NA}} \quad (16)$$

$$\text{Resolution (R)} = \frac{0.47\lambda}{\text{NA}} \quad (17)$$

Rayleigh criterion signifies the limitation of perfect imaging influenced by Fraunhofer diffraction, and defined as the principle maximum of the diffraction pattern of one falling on the first minimum of that of the other. However, Sparrow criterion also points out the same concept but in another condition that both central maximum and the minimum in between just coincide, as shown in Figure 2-6. Although Sparrow criterion provides a stricter but accurate condition, in Rayleigh's own words: "*This rule is convenient on account of its simplicity and it is sufficiently accurate in view of the necessary uncertainty as to what exactly is meant by resolution.*" [38].

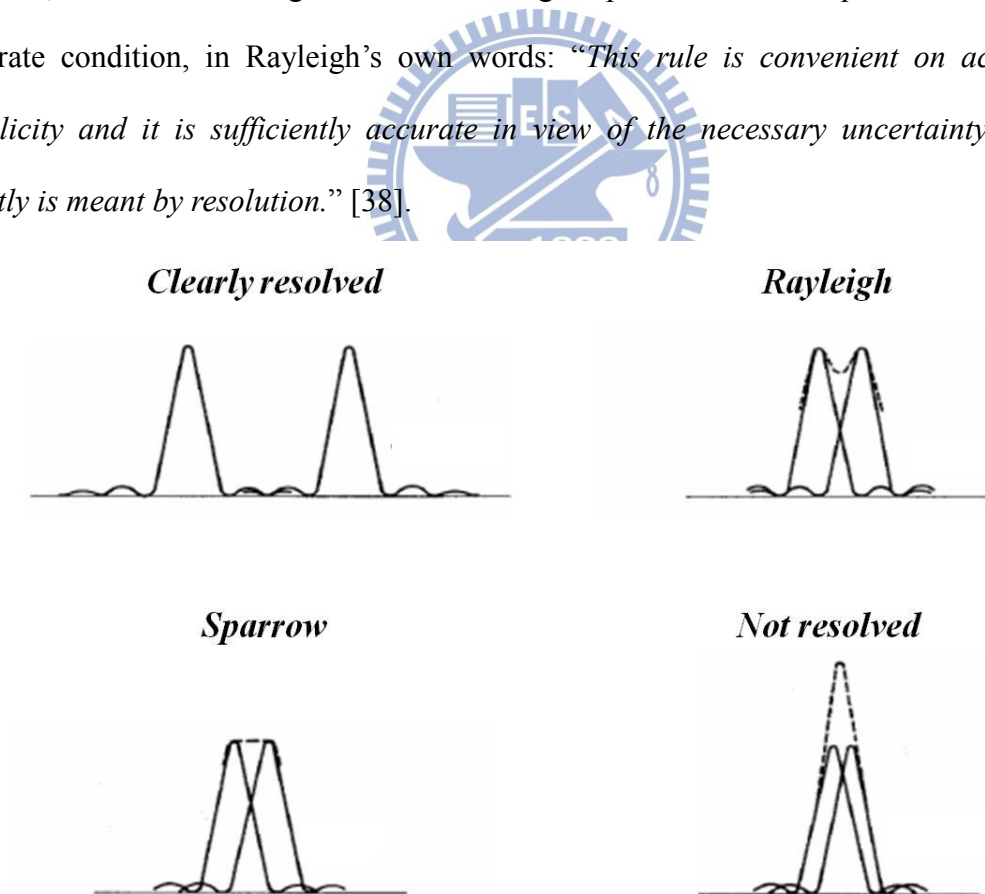


Figure 2-6 Rayleigh and Sparrow criteria for two overlapping diffraction patterns

When it comes to **NA**, **f-number (f/#)**, or focal ratio, resorts to the same characteristic of an image system. The illumination (power per unit area) at image side is inverse proportional to image size, and the image area is proportional to the square of focal length according to the Newtonian form of thin lens equation. Thus, the ratio of image size and aperture size is a quantity of the relative illumination and the square root of this ratio is called relative aperture, f-number, given by [36]:

$$f/\# = \text{efl/clear aperture} = f/D \quad (18)$$

To confirm the equivalence of **NA** and **f/#**, we reconsider the Rayleigh criterion. The intensity distribution **E** of the diffraction of a point source is governed by Bessel function of the first kind, where **E₀** is the central maximum [39].

$$E = E_0 \left(\frac{2J_1(\sigma)}{\sigma} \right) \quad (19)$$

Hence the minimum resolvable separation **R** is

$$R \approx f \Delta\theta = f \frac{1.22 \lambda}{D} = 1.22 \lambda (f/\#) \quad (20)$$

Comparing **R** in terms of **NA** and **f/#**, it is clear that

$$f/\# = \frac{1}{2NA} \quad (21)$$

The two quantities are related for aplanatic systems with infinite object distances; in other words, this equation only holds when the subtended angle is small enough.

Depth of focus (DOF) is also an inevitable express with respect to microscopy because **DOF** is referring to a longitudinal amount within which the imaging is considered clear. In contrast, **depth of field** indicates the range in object space that all object points can be imaged with acceptable sharpness. Although there are two definition of **DOF**, the slight difference is whether the depth of field is symmetric about some reference plane along the optical axis. Figure 2-7 shows the depth of field and depth of focus as the colored regions respectively. And by the geometric relation and Gaussian form of thin lens equation, assuming the system is perfect, i.e. no aberration and no diffraction, **DOF**, **δ**, is given by:

$$\text{DOF} = \delta = \pm c (f/\#) = \pm c / 2\text{NA} \quad (22)$$

where c is the circle of confusion limit, or the pixel size of the sensor. The corresponding depth of field is from s_{near} to s_{far} [36]:

$$s_{\text{near}} = \frac{fs(D + c)}{fD - sc} \quad (23)$$

$$s_{\text{far}} = \frac{fs(D - c)}{fD + sc} \quad (24)$$

where D is the diameter of the entrance pupil of the lens, f is the focal length of the lens, and s is the nominal distance at which the system is focused.

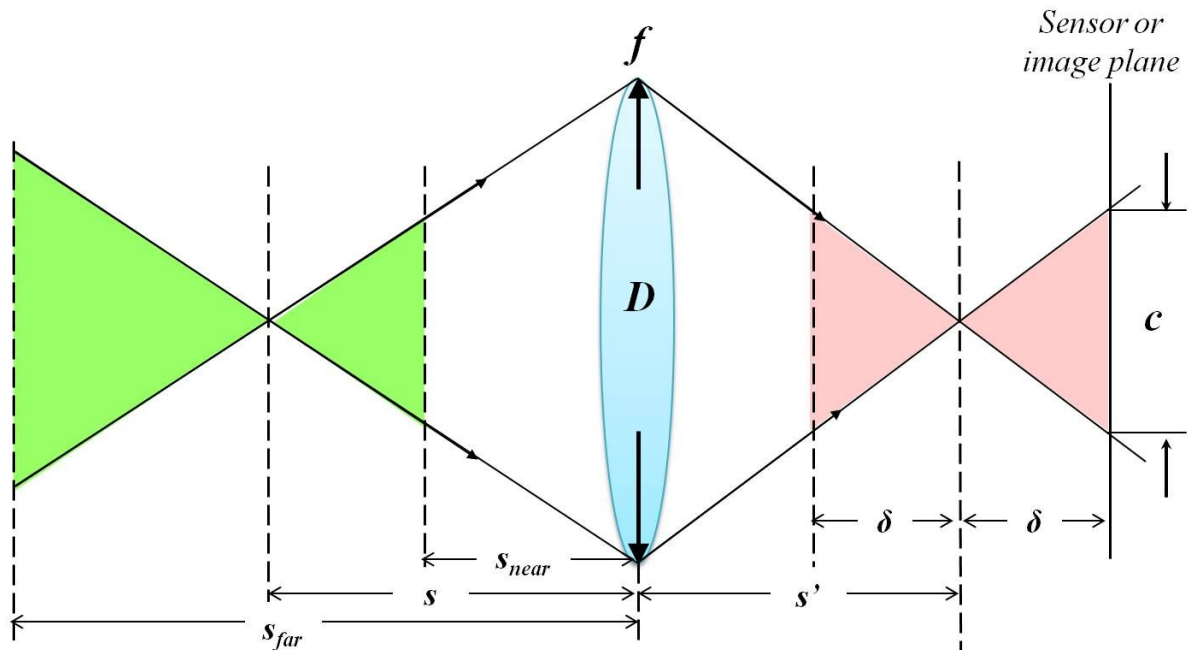


Figure 2-7 Depth of focus and depth of field

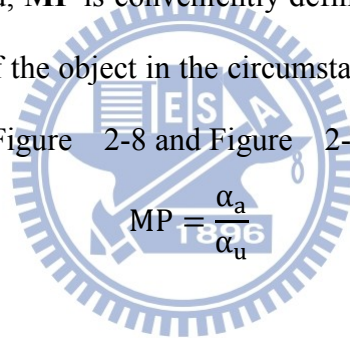
Taking account of D_{far} , **depth of field** will be infinite while

$$s = s_{\text{hyp}} = \frac{-fD}{c} \quad (25)$$

where s_{hyp} is called the **hyperfocal distance** of the system which is a leading distance for fixed-focus cameras [40].

2.2.2 Optical Microscope

The simplest microscope is a magnifying glass which adds refractive power to the eye and provides larger scene that the image seen by unaided eye. Without a doubt, it is desired for the magnifying glass to produce an erect image. Both convex lens and concave lens can form an erect image, but only convex lens create a magnified image. As a result, the object should be placed within the focal length f (e.g., $s_o < f$, s_o is object distance). Concerning the functionality of microscopes, **magnifying power, MP**, or angular magnification is utilized and described as the ratio of the size of the image seen through the optical element/system over the size of the image seen by unaided eye at normal viewing distance d_o , generally taken as 254 mm or 10 inches. Instead, **MP** is conveniently defined as the ratio of the angles made by the chief rays from the top of the object in the circumstance of aided (α_a) and unaided (α_u) eye respectively, as depicted in Figure 2-8 and Figure 2-9..



$$MP = \frac{\alpha_a}{\alpha_u}$$

(26)

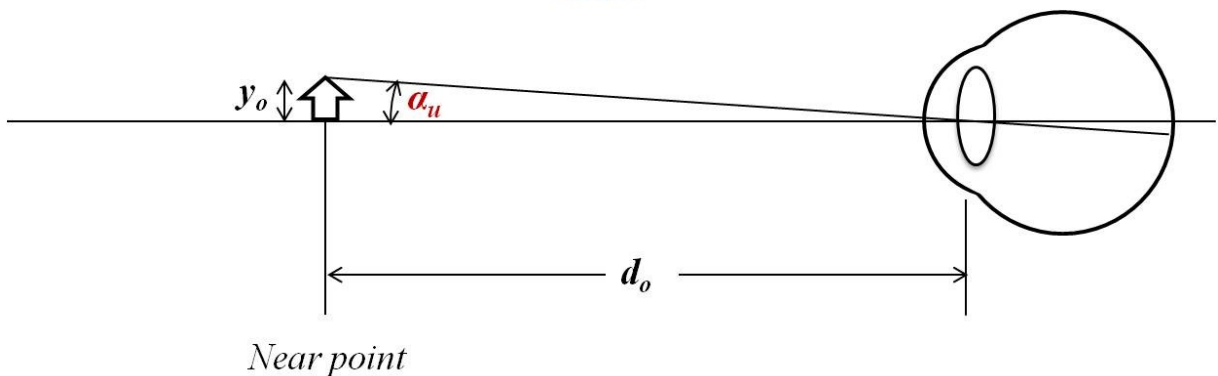


Figure 2-8 An unaided view of an arrow object

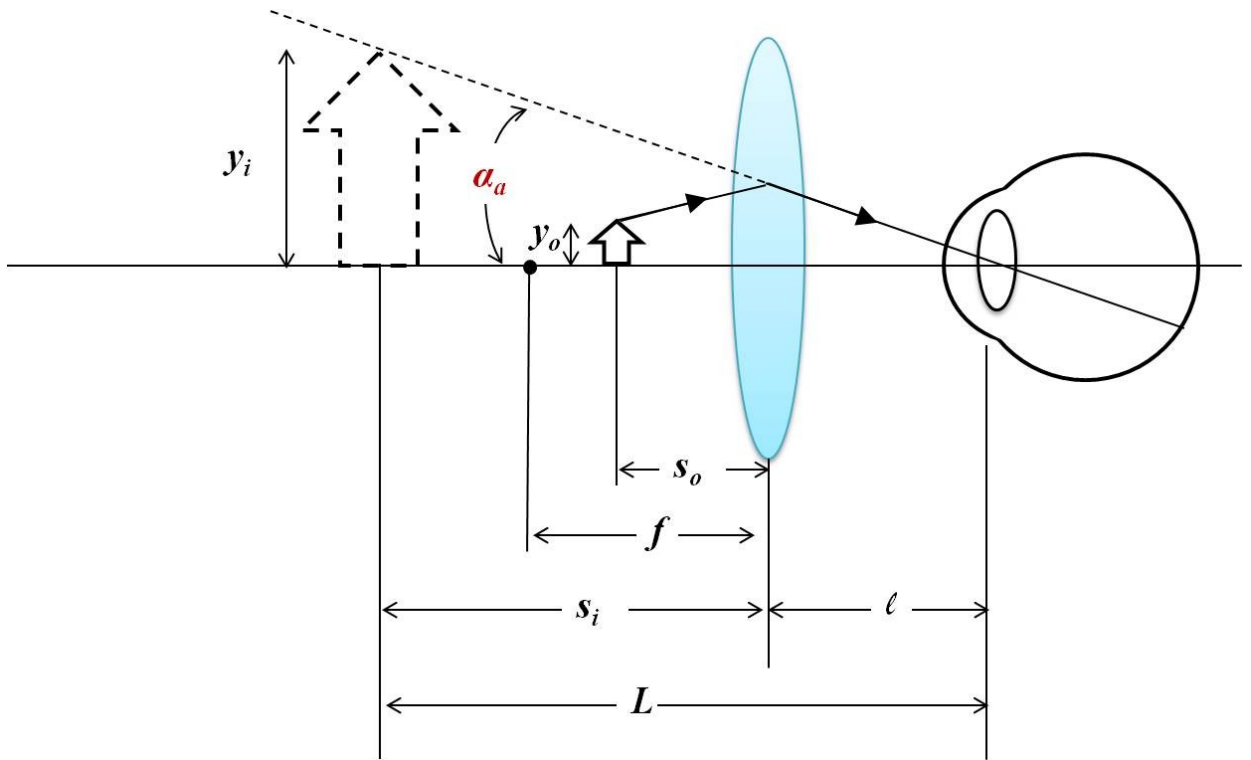


Figure 2-9 An aided view through a magnifying glass

Due to paraxial approximation, $\alpha_a \approx y_i/L$ and $\alpha_u \approx y_o/d_o$, hence

$$MP = \frac{y_i d_o}{y_o L} \quad (27)$$

Besides, grounding on transverse magnification relation and Gaussian Lens Formula, **MP** becomes

$$MP = \frac{d_o}{L} \left[1 + \frac{(L - \ell)}{f} \right] \quad (28)$$

For most common situation that the object is positioned at the focal point, the virtual image is at infinity correspondingly and for all practical values of ℓ , **MP** results in

$$MP = \frac{d_o}{f} \quad (29)$$

It is a pleasing feature that parallel rays procure the relaxed and unaccommodated vision. Nevertheless, **MP** for simplest magnifiers is limited 2X or 3X owing to aberrations, so other more complicate magnifiers are designed up to 20X of **MP**.

In order to provide higher **MP**, compound microscope allegedly invented by H Janssen

and his son Z. Janssen [41] combines two optical units: eyepiece and objective. As implied by the names, eyepiece is a visual optical instrument to adjust a comfortable viewing range and expand the image further while objective is closest to the object and frequently serves as the aperture stop and entrance pupil of the system, as illustrated in Figure 2-10.

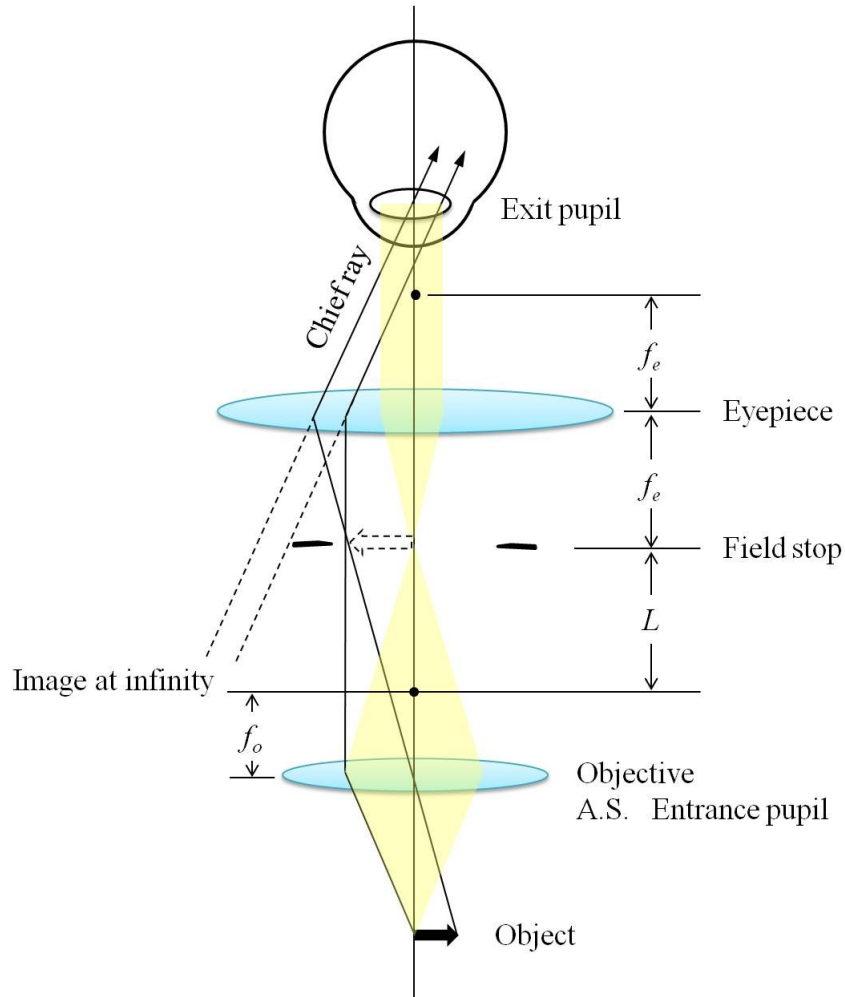


Figure 2-10 A rudimentary compound microscope

Thus, the total **MP** of the entire system is the product of the transverse linear magnification of the object, M_{To} , and the angular magnification of the eyepiece, M_{Ae} .

$$MP = M_{To}M_{Ae} \quad (30)$$

Generally speaking, tube length, denoted by L , is standardized as a constant, 160 mm. If the focal lengths of eyepiece and objective are the interested variables, the total **MP** can be formulated as

$$MP = \left(-\frac{L}{f_o}\right) \left(\frac{d_o}{f_e}\right) \quad (31)$$

where d_o is the standard near point.

Last but not least, angular **field of view** in image side is a vital factor to specify the extent of the largest object that can be viewed. The main component dominating the field of view is the field stop, strictly speaking, is the exit window. The image of the field stop formed by the optical elements following it is called exit window. The cone angle subtended at the center of the exit pupil by the periphery of the exit window is said to be the angular field of view [38].

As a matter of fact, there are still many components needed to structure a compound microscope as shown in Figure 2-11.[41] For example, condenser is a lens designed to concentrate the light onto the specimen. Nonetheless, the key character has been demonstrated and elaborated above.



1. *Eyepiece*
2. *Objective turret, revolver, or revolving nose piece*
3. *Objective lenses*
4. *Coarse adjustment*
5. *Fine adjustment*
6. *Stage*
7. *Light source*
8. *Diaphragm and condenser*
9. *Mechanical stage*

Figure 2-11 Basic optical transmission microscope and its elements

2.3 High Dynamic Range Imaging

Dynamic range is the ratio between the largest and smallest values of a changeable quantity. Humans have high dynamic range (**HDR**) in sight and hearing. People can see the objects under weak moonlight or under bright sunlight. This dynamic range is about 90 dB. However, they cannot achieve the perception in both of the extreme cases at the same time and it takes time to adjust between different visual or hearing situations. In practice, **HDR** data require more space to record in audio or video. Hence, some tricks are used to accomplish **HDR** with narrow recorded dynamic range data. For instance, program makers don't use cue of brightness to display nighttime or daytime scenes. Instead, they utilize duller colors and blue lighting to imitate the way that human eyes perceive at low light levels. [42]

In photography, the conventional format of digital images is bmp or jpg (jpeg) which generally use 24 bits for each pixel, and each pixel contains 3 primary colors, so the range of gray levels is from 0 to 255. In other words, the contrast ratio of the images is 256:1 and it is sufficient for most of the scenes. However, if the scene is exposed under the sunlight, it would extend the contrast to 50,000:1.

HDR images possess all the information under different exposure and have wide range of luminance information, so it uses more than 12 bits per channel to cover the large luminance range between the highlights and the shadows. Typically, **HDR** images can be created by composing the images under different exposure time. Imaging technology makes it possible to capture and storage of **HDR** images; nevertheless, the output limitation of common displays has not followed the advances. Therefore, several algorithms are designed to adjust the range of luminance of the real world so that **HDR** images can be displayed on the devices with lower dynamic range. [43]

2.3.1 High Dynamic Range Imaging Rendering Algorithm

HDR image rendering algorithms can be categorized into two types: global operators and local operators. Global operators apply same processing over one image based on the image content while local operators use different mapping methods according to spatially localized content. Notwithstanding global operators benefit faster computation and easier to implement, local operators allow for larger dynamic range compression. The following are brief introduction to these operators.

Sigmoidal Transformation: Sigmoid contrast enhancement function $S(t)$ derived from a discrete cumulative normal function is utilized to rescale the lightness for gamut mapping. This method was presented by Braun in 1999. Afterwards, this method is modified to compress the HDR images by the logarithm of luminance.

$$S(t) = \frac{1}{1 + e^{-t}} \quad (32)$$

Histogram Adjustment: By incorporating the human visual models of glare, spatial acuity and color sensitivity effects, the histogram of luminance is modified to reproduce the imperfections in human vision, which was proposed by Ward in 1997.



Figure 2-12 Dodging and burning effect

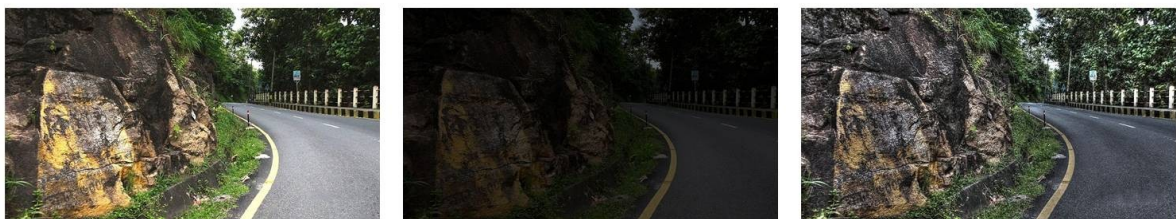
Photographic Reproduction: Different luminance mapping of highlight and shadow region is applied to simulate the dodging-and-burning effect in traditional photography. As Figure 2-12 illustrated, Dodging decreases the exposure to make the film negative brighter

while burning increases the exposure to make it darker. [44] This tone mapping techniques was presented by Reinhard et al. in 2002.

Bilateral Filtering Technique: An image is decomposed by an edge-preserving spatial filter into base layer and detail layer. Base layer contains large scale of variations. The overall brightness and base contrast is compress and subsequently two layers are combined into final image. This technique proposed by Durand and Dorsey in 2002 reduce the overall contrast but maintain the local details in the image.

Local Eye Adaption: The Naka-Rushton equation is modified to predict the response of cones and rods. According to the S-shaped response function, the luminance channel is compressed. This local-eye adaption method is presented by Ledda et al. in 2004.

The aforementioned algorithms [44] [45] are just few of them, but the core idea is trying adjust the range of luminance in the image with our conventional displays due to the limited contrast ratio. In brief, **HDR** images provide more realistic visualization of the real word as what people perceive. If one day, researchers surmount the limited capability of displays and our knowledge to human visual system, more robust models and operators can be utilized to improve the perceptual accuracy. Finally, Figure 2-13 shows an example of HDR image rendering by two images with different captured intensity. [47]



(a)

(b)

(c)

Figure 2-13 Example of high dynamic range image by tone mapping
(a)+2 stop (b)-2 stop (c) HDR image

Chapter 3

Structure and Algorithms

Binocular vision is the foundation of 3D capturing with lens array. We utilize the disparity of elemental images to render the depth information. However, the corresponding problem is always an issue for stereo camera. Lens array in our High Dynamic Depth Range (HDDR) system is thus modified with different focal arrangement and the configuration of temporal and spatial systems will be illustrated in the first part of this chapter. Besides, depth information is extracted by the Depth Estimation Reference Software (DERS) and the post image processing, Depth map Fusion from Edge Exploring Thresholding (DFEET), is carried out in order to fuse the depth maps together. In the other part of this chapter, we will carefully elaborate on each step in DERS and DFEET. Finally, the limitation of our algorithm will be discussed as well.

3.1 3D Image Capturing with Lens Array

In conventional microscopes, depth of field decrease as magnification increase, so it needs to adjust the focal plane to clearly observe the specimen. Furthermore, some techniques of 3D image capturing are restricted in near field due to the shallow depth of field as well. Blurred images would bring about the matching error as computing the disparity. Even for the light field camera, the reversibility of light rays is the first hypothesis. As a result, when a source point diverges to be a spot, the light field function could not be a one-to-one and onto function anymore, so the integral fails to render back. To eliminate the out of focus issue, variable focal regions can be utilized by the lens array. Concerning the resolution of elemental images, the fewer lenslets are used, the more depth information can be acquired.

To minimize the number of lenslets, each depth of field should be well arranged. According to concept and equation described in Chapter 2, depth of field evolves from depth

of focus with the consideration of longitudinal magnification. For the sake of convenient formulization, we only cogitate upon symmetrical depth of field as

$$r = s - s_{\text{near}} = \frac{-fsc - s^2c}{fD - sc} \quad (33)$$

Objects within twice of r will be clear imaged and other parameters are identical in Chapter 2. By means of the aforementioned equation, different depth ranges of interest can be designed as shown in Figure 3-1. The concept of extended depth of field is that one depth of field overlaps with the closest adjacent depth of field. Ideally, the objects falling out of the depth of field are out of focus, so we can render part of depth in the overlapping depth of field; then stack them altogether. Hence on the basis of reducing the wasting depth of field, the total range of clear imaging is elongated from depth of field of s_1 to wider than that of s_3 , for example. And this concept is similar to **High Dynamic Range (HDR)** in photography, but in depth map, we use depth as the estimated quantity. So we come up with the term: **High Dynamic Depth Range (HDDR)** to stand for the wide range of depth rendering.

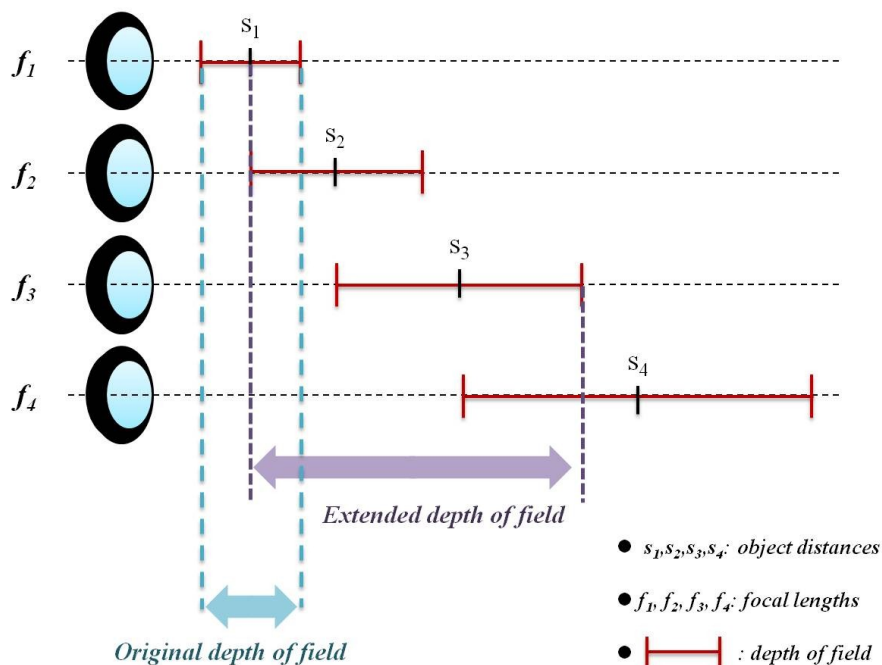


Figure 3-1 Scheme of extended depth of field

However, in point of image processing, the fuzziness is requested more strictly. As a

consequence, we had better transfer from the idea of depth of field to that of Modulation Transfer Function (MTF) [48]. MTF is the frequency response of an optical system, so it governs how sharp the edge is or the contrast of the image since edge is regarded as a high frequency component. The idea of MTF also reveals a fact that the number of different focal lengths can be reduced. So we begin with the simplest case: two depth of field. Note that we use the term depth of field for the idea that images are “in focus” within the layer. Due to the occlusion issue, three elemental images are required to render one depth map. As a result, lens array should comprise at least six lenslets with two focal positions. Both spatial-multiplexed and temporal-multiplexed type can fulfill this idea of HDDR as illustrated in Figure 3-2 and Figure 3-3. Subsequently, we will render two depth maps that are partially well-defined and finally we fuse them together into HDDR depth map.

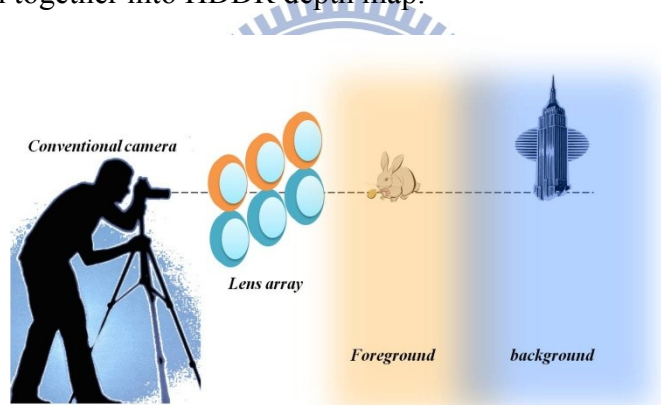


Figure 3-2 Spatial HDDR system with 2 DOF



Figure 3-3 Temporal HDDR system with 2 DOF

Of course these two methods are not totally equivalent because of the deviation in height

between two depth maps. This deviation will lead to distortion of objects because they are captured from different perspectives. When we carry out fusion of two depth maps into HDDR depth map, it is more difficult to reconstruct the correct shape of the objects. However, this issue will be mitigated as the pitch of lens array is getting smaller.

In addition, the result of integrating three depth of field will be demonstrated to prove that this HDDR concept can be applied to a wider range as long as the number of lenslets is increased. As the Figure 3-4 shows, every row of lenslets contributes one depth of field; then depth of field can be greatly enhanced by stacking each layers.

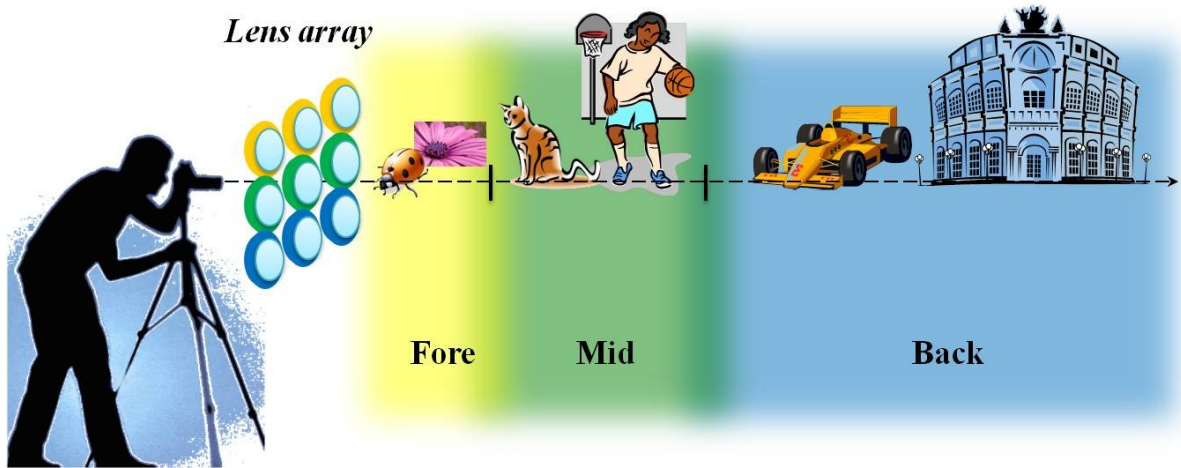


Figure 3-4 HDDR system with 3 DOF

However, there is an ambiguity to distinguish the demarcation of foreground and background. Likewise, this ambiguity stems from the degree of fuzziness the software can tolerate. The more the objects blur, the less accurate the matching will be, because of lacking sufficient feature points. Besides patterned ground and terminal wall are applied to make sure the feature point is enough for the software to find the corresponding points and then compute the disparity. It should be reminded that we don't use large f-number to increase the range of depth of field owing to the low light efficiency. If we want to maintain the captured intensity, dimming environment or objects with quick motion will lead to a dilemma of exposure time.

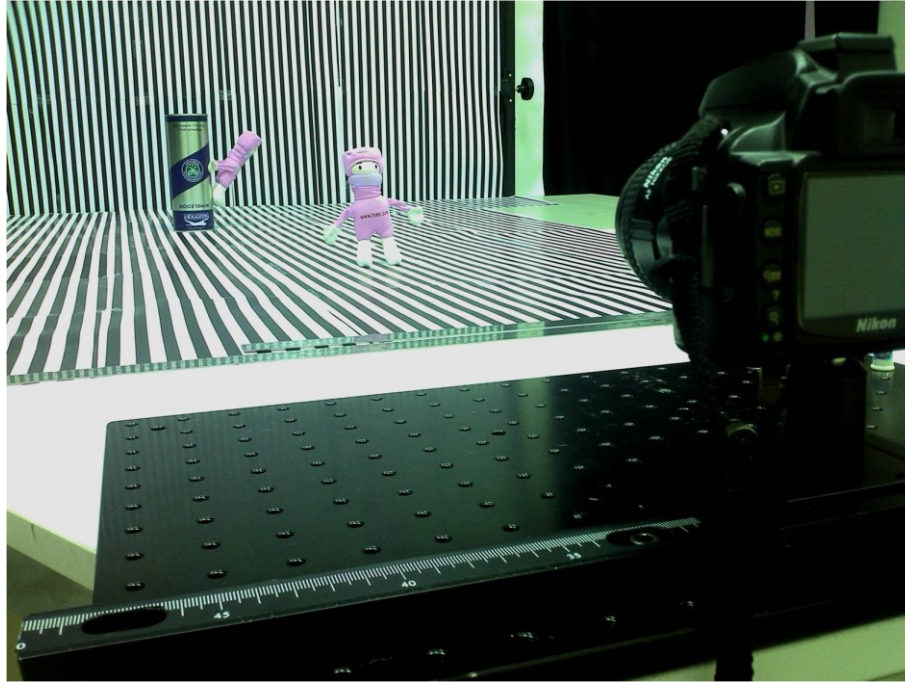


Figure 3-5 Experiment setup

	2 depth of field	3 depth of field
Camera	Nikon D60	Nikon D60
Lens array	1x3 or 2x3	1x3
System	Temporal/Spatial HDDR	Temporal HDDR
f-number	F/2, F/2.8, F/22	F/2.8, F/22
# of elemental images	6	9
resolution of elemental images	1200x800	1200x800
Pitch	1cm or 5 cm	1 cm
Object distance	87, 150cm or 110, 208, 235cm	35, 76, 152 cm

Table 3-1 Experimental Parameters

Last but not least, the experiment setup and parameters are shown in Figure 3-5 and Table 3-1 respectively. In the experiment, we utilize moving camera to simulate the HDDR system. Although we did not use “real” lens array in our experiment, the moving pitch of 5cm

between the elemental images was determined by the size of our camera lens. As a result, it would be feasible to use lens array in the future. Figure 3-6 illustrates the analogy of lens array system and moving camera. We can regard the lens and the main lens as a single optical system, so the effective focal length can be calculated. And the possibility of using moving camera is proved if we carefully adjust the focal length and the moving pitch. However, our camera lens is a prime lens, so we cannot change the focal length but change the image distance to sweep the focal plane. Consequently, we should modify the magnification as we use real lens array in the future. As long as the resolution of elemental images is sufficient, we can determine the factor for resizing according to the depth because the reserved range of conventional depth maps can be first acquired via edge information. In addition, when we use coplanar lens array with different focal lengths, the image planes of the lenslets might be different. Therefore, in Figure 3-7, our target is that the depth of field of our camera should cover the variation of the image planes in order to clearly capture every elemental image. And for the fear that the elemental images would overlap with each other, we have to cautiously arrange the focal lengths and the field of view of each lenslet. To conclude, there are some differences between using lens array and moving camera, but they do not contradict our original concept of HDDR system.

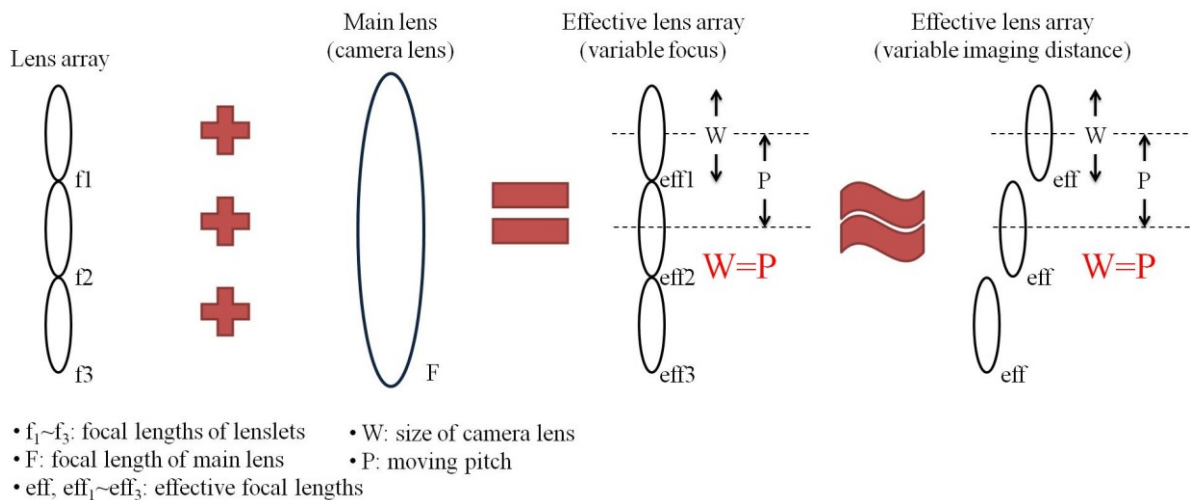


Figure 3-6 Effective lens design

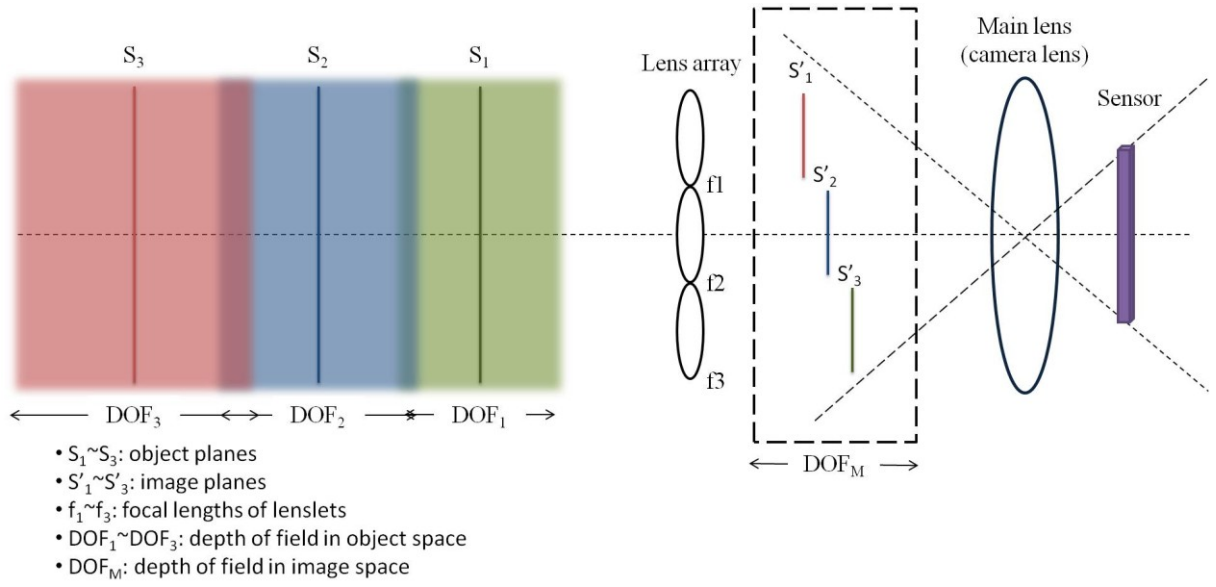
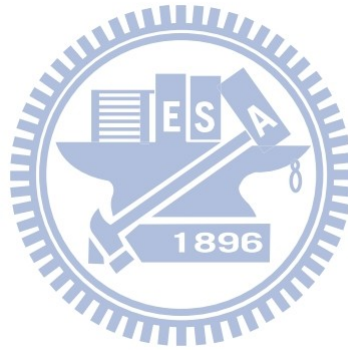


Figure 3-7 Overall imaging system with lens array



3.2 Algorithm

“In the logician’s voice:

‘an algorithm is a finite procedure, written in a fixed symbolic vocabulary, governed by precise instructions, moving in discrete steps, 1, 2, 3, ..., whose execution requires no insight cleverness, intuition, intelligence, or perspicuity, and that sooner or later comes to an end.’”

— *“The Advent of Algorithm” by David Berlinski, 2000*

After capturing the elemental images, we have to visualize the depth information. In terms of the transmission of 3D TV, the most straightforward and common way is to use a depth map. Based on a monoscopic video, the corresponding depth map can be utilized to synthesize the stereo image pairs for more virtual view of 3D scene. This technique is denominated as Depth-Image-Based Rendering (DIBR). [49] To generate a HDDR depth map, the overall flow is shown in Figure 3-8.

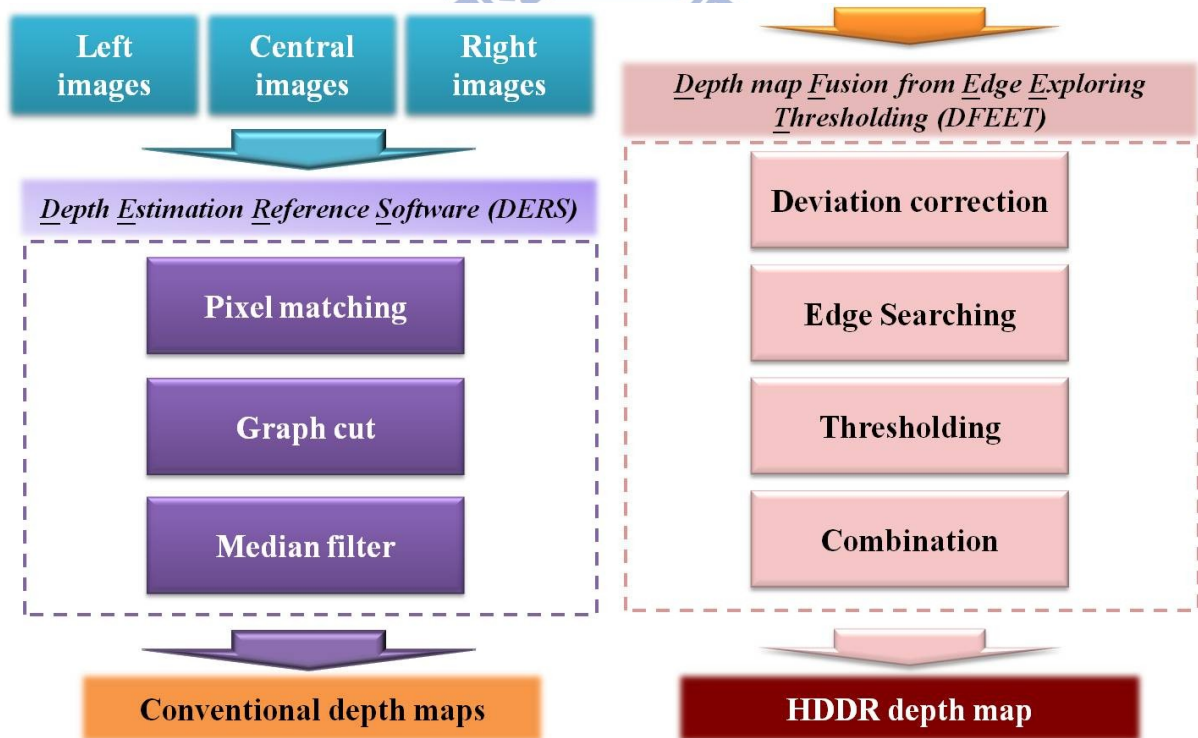


Figure 3-8 Flow chart of algorithm

Our algorithm can be roughly divided into two parts. One part is using stereo matching software to generate conventional depth maps. The other part is the fusion of the conventional depth maps. The following two sections will disclose the details of the software and how to fuse all depth map together.

3.2.1 Depth Estimation Reference Software (DERS)

Moving Picture Experts Group (MPEG) is an authoritative group in charge of developing global standards of compression, decompression, and coded representation of pictures and audio [50]. On September 24th 2009, a latest version of the matching tool, called Depth Estimation Reference System (DERS) is released on London meeting. The latest version, DERS 5.0, introduced a technique of soft segmentation matching because one of the most crucial elements in depth estimation is matching the corresponding points between the stereo images. In first version of DERS, direct pixel matching approach is vulnerable to noise. As a result, a weighted comparison mask is utilized in DERS 5.0 to describe the significance of pixels neighboring with processed pixels. [51] Moreover, DERS is different from other depth map generators because it uses three input views instead of two and it is able to perform depth estimation for image sequences. Three input views stand for left, central, right views which are equally spaced along the horizontal baseline. The benefit of introducing third view is that the occlusion issue is somewhat reduced.

Generally speaking, the disparity estimation in DERS includes the following steps [52]:

a. Image segmentation (optional)

Three methods including *mean shift*, *pyramid segmentation*, and *K means clustering*, can be chosen but the bandwidth parameters are fixed in the source code.

b. Pixel/block matching

Two options for matching: 1.) the simplest approach is pixel matching, which

compares the intensity differences pixel-wise. 2) The other approach is block matching, which uses a 3-by-3 window with adaptive weights to compute the cost function. The weights are given as following formula:

$$W(P, P') = \exp \left(-\frac{|I(P) - I(P')|}{\gamma_c} - \frac{|P - P'|}{\gamma_d} \right) \quad (34)$$

where

P = center point in processed frame,

P' = processed point in processed frame,

$W(P, P')$ = soft-segmentation mask around center point P ,

$I(P)$ = intensity of image at point P ,

$|P - P'|$ = Euclidian distance between P and P' ,

γ_c = color similarity parameter,

γ_d = distance similarity parameter.

c. Cost adjustment for temporal enhancement (optional)

This function is especially useful as dealing with image sequences, which updates the cost function by the block motion detection. The data cost of static blocks is set as zero to encourage same disparity will be selected again during graph cut optimization.

d. Disparity computation using graph cuts optimization

Graph cut is one of most common optimization methods in stereo correspondence work and it bases on two strategies: α , β -swap moves and α -expansion moves. The key idea of graph cut is trying to minimize the energy function $E(d)$.

$$E(d) = E_{\text{data}}(d) + \lambda E_{\text{smooth}}(d) \quad (35)$$

where λ is to adjust the smoothness and $d=d(x,y)$ is the disparity map.

While judging the similarity of corresponding points, dissimilarity is equivalent. So, we define a cost function $C(x, y, d)$ in disparity space to return the value of dissimilarity. And $E_{\text{data}}(d)$ is the summation of the cost.

$$E_{\text{data}}(d) = \sum C(x, y, d(x, y)) \quad (36)$$

As for the smoothness term, it is a function or called penalty $\rho(d, I)$ often depending on differences in disparity and intensity of neighboring pixels. $\rho(d, I)$ increases with disparity difference but reduces as the discontinuities locate on the color edges. Then $E_{\text{smooth}}(d)$ is the summation of the penalty.

$$E_{\text{smooth}}(d) = \sum \rho(d(x, y) - d(x + 1, y), I(x, y) - I(x + 1, y)) + \rho(d(x, y) - d(x, y + 1), I(x, y) - I(x, y + 1)) \quad (37)$$

In DERS, α -expansion moves approach is used while doing graph cut optimization. As Figure 3-9 depicted, pixels with any labels in original labeling might be assigned with the new label α . By means of appropriate grouping the pixels, the energy function will be minimized.

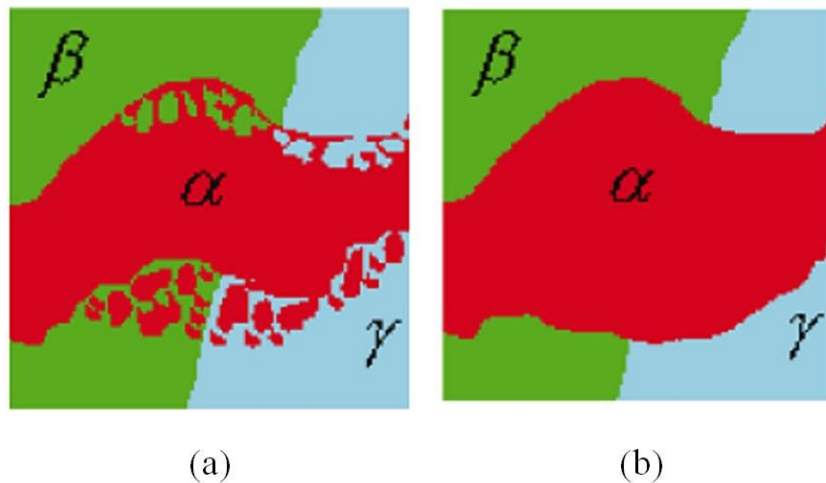


Figure 3-9 Example of results from moves in graph cut algorithm.(a) Original pixel labeling (b) after an α -expansion move

- e. Refinement of created disparity maps by using plane fitting (optional)

This step is triggered as image segmentation is activated. Least squares method is

utilized to refine the previous results of segmentation.

- f. Post processing: 3-by-3 median filter

Final step is to further reduce the noise via median filter. However, the size is predetermined.

3.2.2 Depth Map Fusion from Edge Exploring Thresholding (DFEET)

In DFEET, four steps are executed: 1.) deviation correction 2.) edge searching 3.) thresholding 4.) Combination.

First of all, when we use spatial HDDR system, two depth maps are not rendered from the identical height. They are similar to elemental images that include disparity. Therefore, we have to adjust one depth map to the same height with the other depth map. For example, if we use the depth map from the nether position as the model, we have shift the upper depth map downward to keep the positions of objects are the same. Figure 3-10 illustrates an example of deviation correction. Red image and green image are the original image captured from upper position and shifted image respectively. It should be noted that deviation correction is carried out after the conventional depth map is generated. Hence, both of the color image and depth map should undergo shifting.

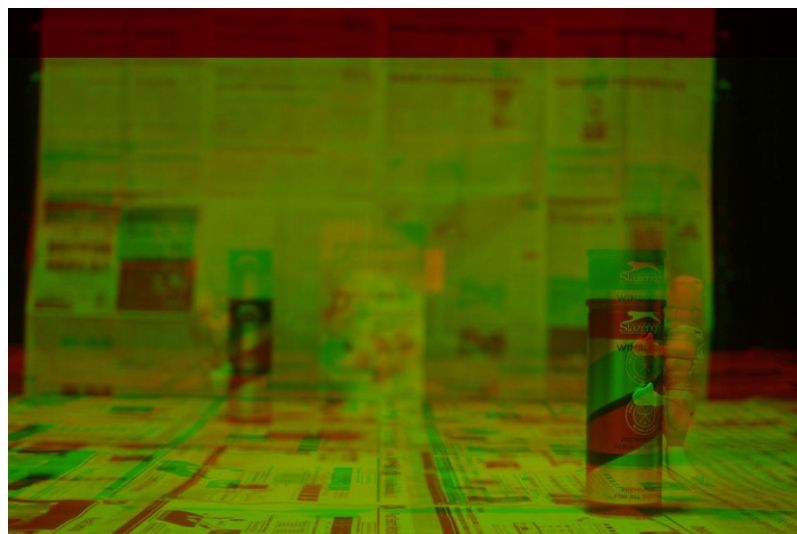


Figure 3-10 Deviation correction

It goes without saying that if temporal HDDR system is applied, there's no need to do the deviation correction, i.e. disparity is zero.

Secondly, edge is often regarded as the indication to judge the things whether they are in focus. Consequently, we apply edge filter to analogize the region of depth of field. When it comes to edge filter, high pass filter is the basis of edge detection. However, high pass filter such as Laplacian operator is sensitive to noise, so it requires noise suppression beforehand. Fortunately, Marr and Hildreth proposed Laplacian of Gaussian (LoG) method to take care of these two considerations. We can create a mask by sampling the following equation. [53]

$$\nabla^2 G(x,y) = \left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (38)$$

where $G(x,y)$ is a Gaussian function. As the name implies, we perform Gaussian smoothing as well as Laplacian sharpening. LoG is one of the common approaches used in edge detection and it is easy to implement with acceptable accuracy, so we use it in our algorithm. After extracting the edges of the elemental images, we apply another identity matrix to convolve with edge image in order to find a representative focal point.

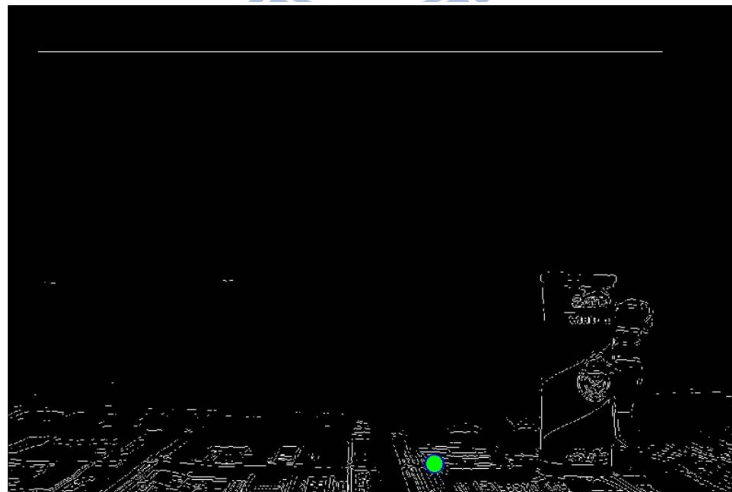


Figure 3-11 Representative point of a deviation-corrected elemental image

In Figure 3-11, in addition to the line above due to the deviation correction, the green point implies that this position is full of edge information, so we regard it as the representative focal point. Conceptually, we consider this point as the location of focal plane when we capture the

image but actually it is not always correct. When the content is texture-less or with low contrast, the representative focal point would locate far away from what we expect. This phenomenon will influence the exactness of determining the threshold value in next step. However, there's an ambiguity between two depth of field as mentioned before, so the variation of the threshold value is acceptable. Unless the representative focal points deviate too much, segmentation error of depth map will occur since the boundary is distinct from the region of ambiguity. In other words, ill-defined objects will not be filtered out after thresholding.

With regard to thresholding, once the representative focal point is discovered, we can find the corresponding gray level from the depth map. By the same token, the representative focal point and its corresponding gray level can be detected as well. As a result, if N depth of field are arranged in capturing, N representative focal points will be extracted. Subsequently, N-1 threshold gray value can be decided by averaging the corresponding gray levels of two adjacent representative focal points as illustrated in Figure 3-12. The meaning of thresholding in depth maps is equivalent to separation of each depth of field along the scene. But in the depth map, most of the objects become planar because each of them includes only few gray levels. Therefore, we don't have to worry about the exact position while thresholding.

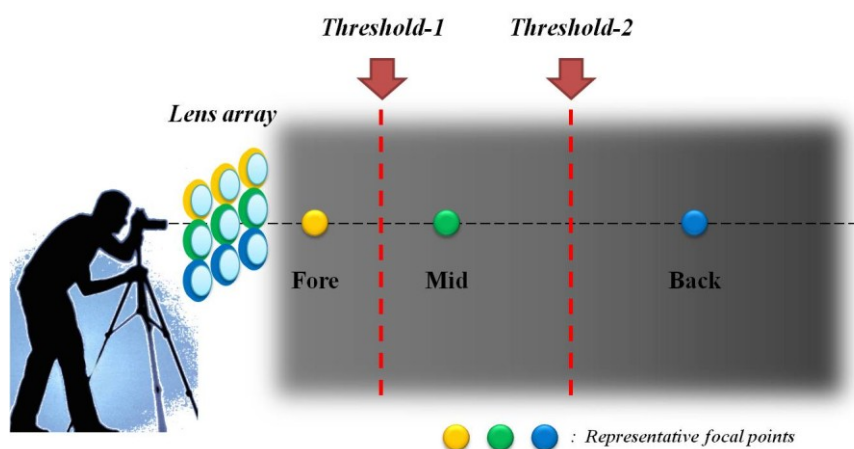


Figure 3-12 Scheme of threshold determination

However, due to the noise or other factors that cause matching error while generating conventional depth maps, the gray level varies inconsistently among the depth maps. Accordingly, each segment cannot be carefully combined. As shown in Figure 3-13, red, green, blue parts stand for three segments and when they are composed together, voids (black) and overlapping (yellow and cyan) will appear between the boundaries of each segment.

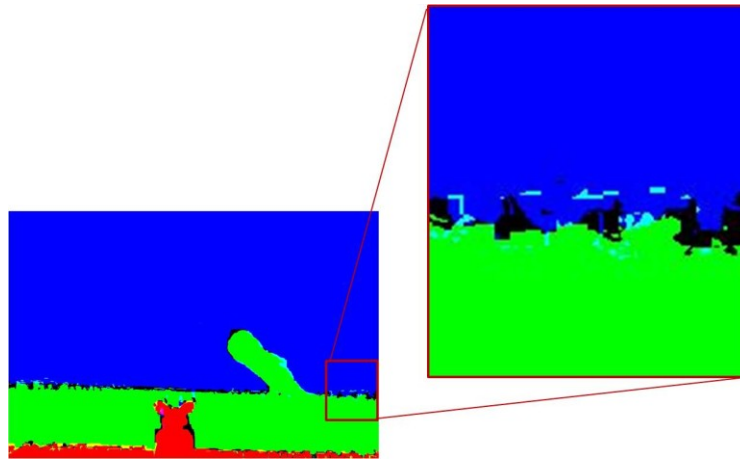


Figure 3-13 Segmentation voids

Hence in the last step, combination, we have to deal with these two problems: voids and overlapping. As for overlapping, the relative large gray level is selected because it is on behalf of the front object point. This concept is accordant with the experience in real world. However, the issue of voids is more difficult to handle because it is a process that makes something out of nothing. Therefore, we have to rely on the side information of the voids. Least smoothing operators will degrade the contrast between the object and its background, we use median filter to reconstruct the voids.

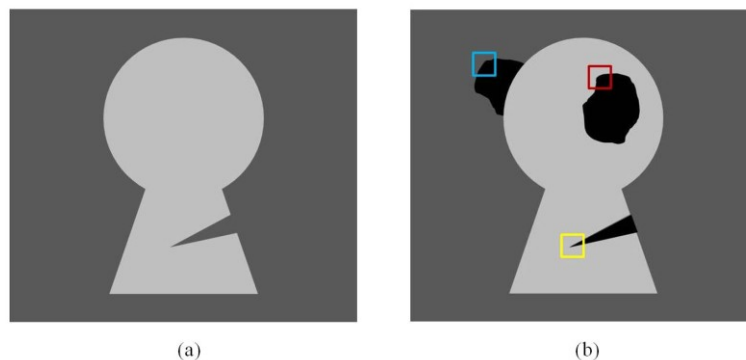
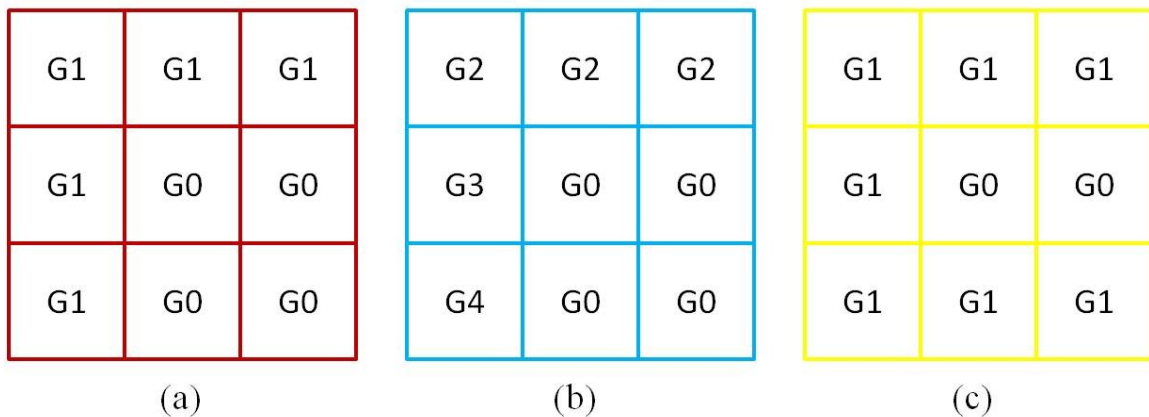


Figure 3-14 Three situations while reconstruction (a) ideal image (b) voided image

There are three situations during reconstruction as the three blocks in Figure 3-14 . Firstly, red block means the pixel of the void is luckily surrounded by its correct object points, so same gray level will be assigned by median filter, as depicted in Figure 3-15 (a). After sorting, the central pixel G0 will be substituted by G1. Likewise, blue block shows the void is encompassed by the similar background point we want to reconstruct. However, the background of depth maps is gradually varying, so reconstruction will lead to small mistakes by median filter, as shown in Figure 3-15 (b). Although the background points G2~G4 are in the majority, G0 will mistakenly be replaced with G2, if the normal direction of camera is parallel to the extension of depth. In other words, G0 should be replaced with G3. Nevertheless, the result will be different with only few gray levels if the background is with no noise, i.e. subtracting G2 from G3 is almost zero. But for the third situation, dramatic errors are caused especially for regions of small aspect ratio. As the example in Figure 3-15 (c), central G0 should be changed with background points, but it is assigned with G1 instead. This error leads to a region growing which makes the shape of the object distorted.



G0: gray level of voids, G1: gray level of specific object, G2~4: gray level of background of depth map

Figure 3-15 Examples of median filtering of three situations of reconstruction (a) object reconstruction (b) background reconstruction (c) error reconstruction

Chapter 4

Experiments and Results

Based on the system and algorithm illustrated in chapter 3, the experimental results will be shown here and it includes four parts. First, we reveal how the image quality will influence the depth map rendering. Subsequently, HDDR depth maps are rendered by stacking two and three depth of field respectively, and compared with the result captured by the largest f-number of our camera (F/22). Moreover, the feasibility of lens array will be verified in the last part via considering the actual pitch of each lenslet. Finally, the characteristics of temporal and spatial HDDR system will be summarized.

4.1 Depth Maps of Under-exposed and Blurred Images

As mentioned in Chapter 3, we don't use large f-number to increase the range of depth of field owing to the low light efficiency. If we want to maintain the captured intensity, dimming environment or objects with quick motion will lead to a dilemma of exposure time. Therefore, we first examine how under-exposed or blurred images would affect the depth maps. Figure 4-1 shows the colorful elemental images captured with large f-number (F/22). Obviously, although the exposure time is only one tenth shorter, the contrast of image decreases dramatically. Compared to our spatial HDDR system, the exposure time of (d)~(f) is even several dozen times longer. Note that (g)~(i) is the adjusted images simply for the fear that image information in (d)~(f) is not clear enough. Due to the insufficient image information of under-exposed image, it is difficult to render a good depth map. As shown in Figure 4-2, all the objects are ill-defined in the depth map. Because the ambient light is not uniform, the result of rear objects is worse than that of front object.

Although blurred effect of out-of-focus images is not exactly equivalent with that of ghost images due to motions, both of them result in soft edges however. If we regard edge as

an important feature, the two cases are similar to some degree. Consequently, we firstly utilize circular disk to generate the defocused images [54]. By rendering the depth map from the defocused image as shown in Figure 4-3, the error trend can be simulated. Larger object will lead to more inaccurate pixels, so the error is calculated from the ratio of the wrong pixels and the size of object and the result is exhibited in Figure 4-4. In experiment, we also use blurred elemental images to testify whether there will be lots of mismatch in the rendered depth map. Three blurred elemental images are shown in Figure 4-5. Although we maintain most of the color information so that human can still roughly define the shape and locations of objects, the fuzziness of objects is far beyond that in color images.

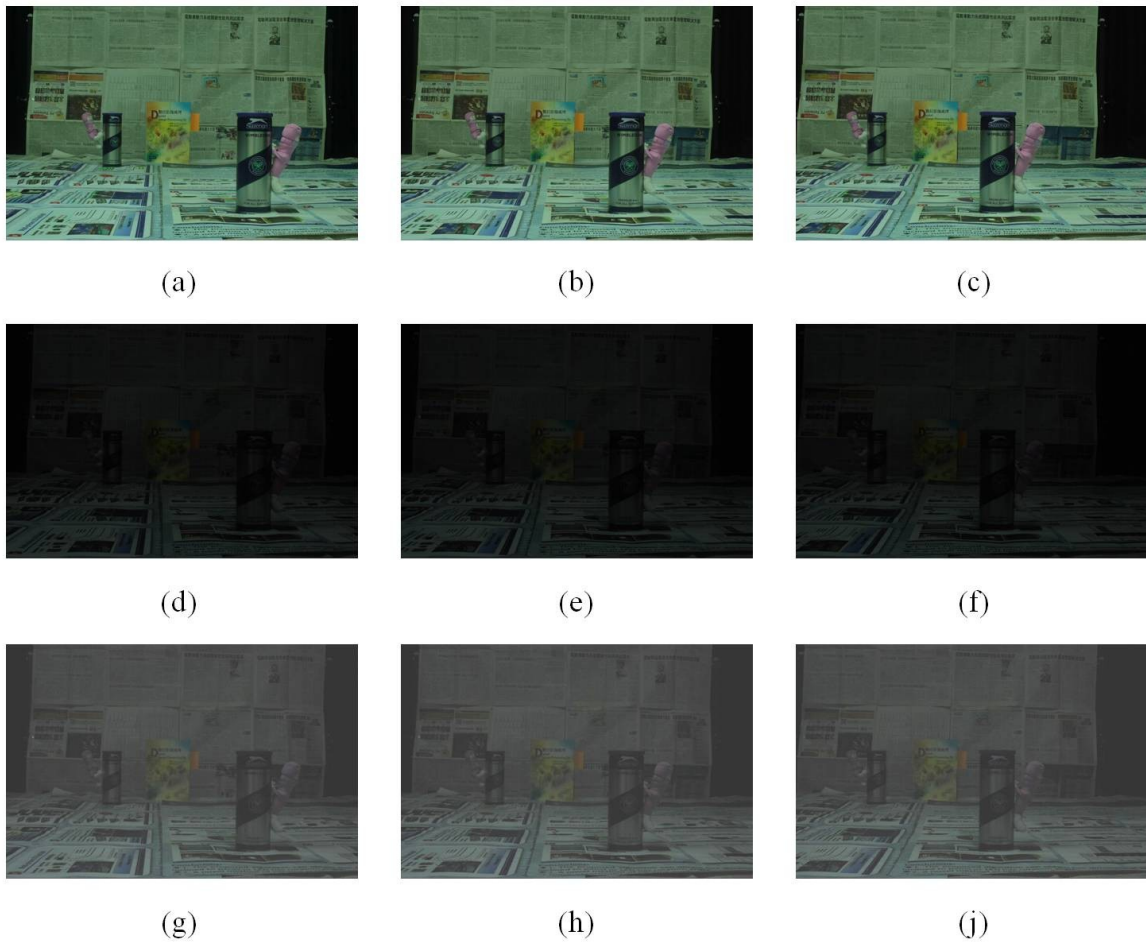


Figure 4-1 Elemental images under F/22 with different exposure time (a)(b)(c) three perspectives with adequate EV, (d)(e)(f) one tenth of adequate EV, (g)(h)(i) adjusted images of (d)(e)(f) respectively

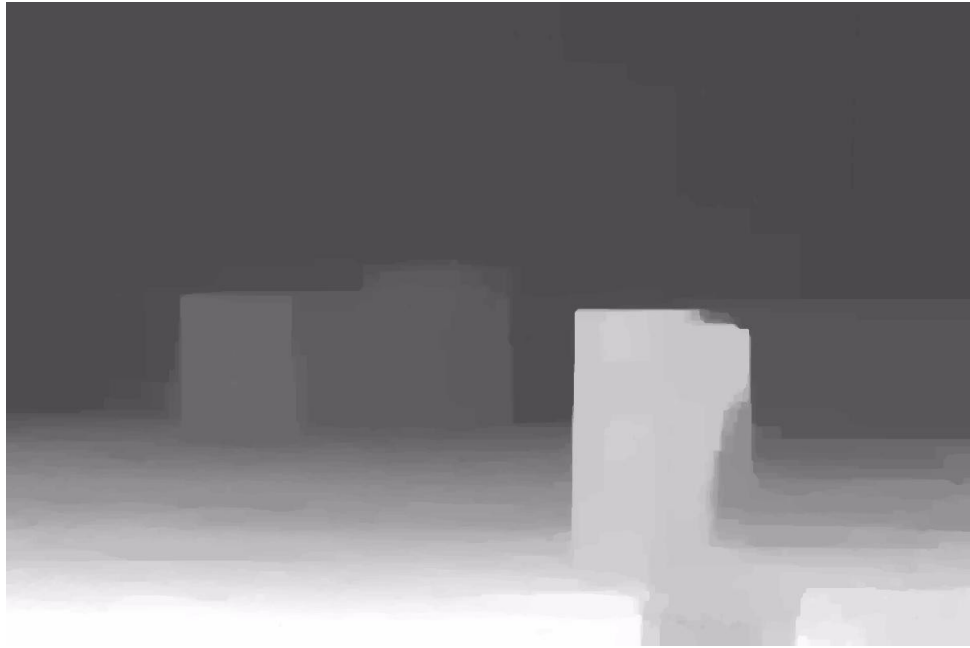


Figure 4-2 Rendered depth map from under-exposed elemental images

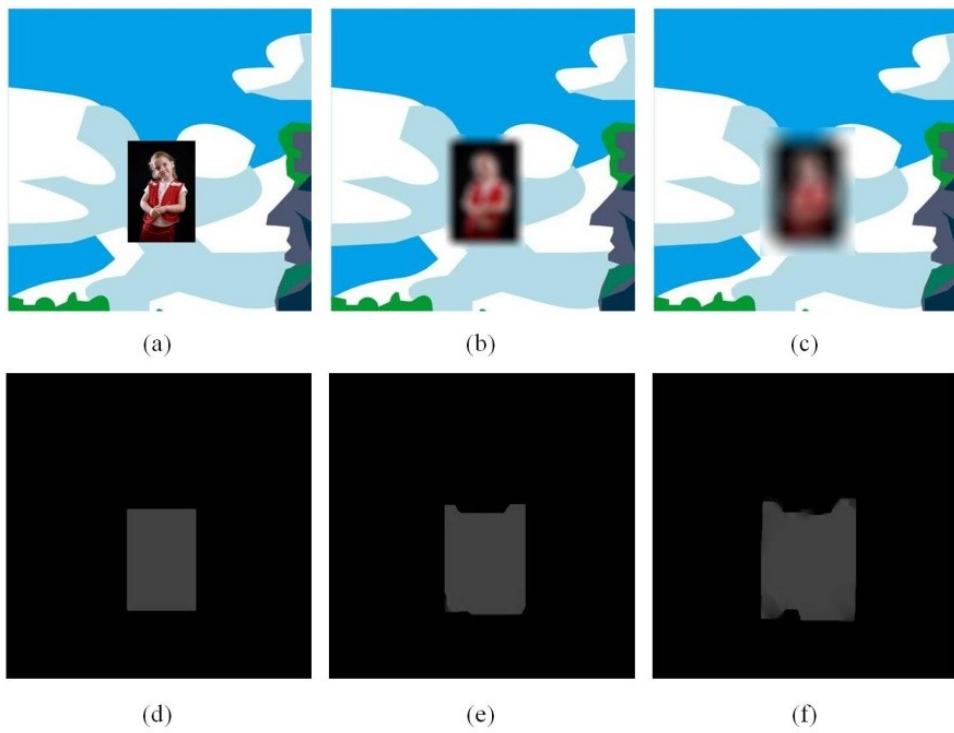


Figure 4-3 Blurred image and its depth map (a)(d) ground truth (b)(e) 21-pixel variance (c)(f) 41-pixel variance

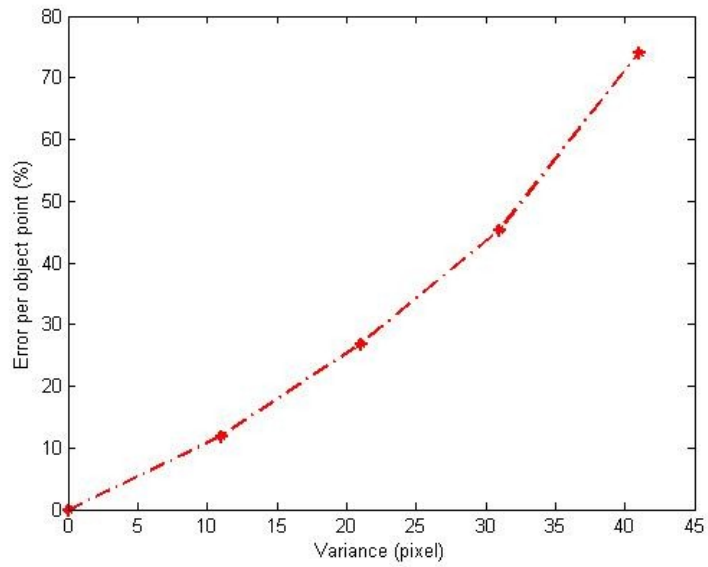


Figure 4-4 Error rate versus the variance of disk

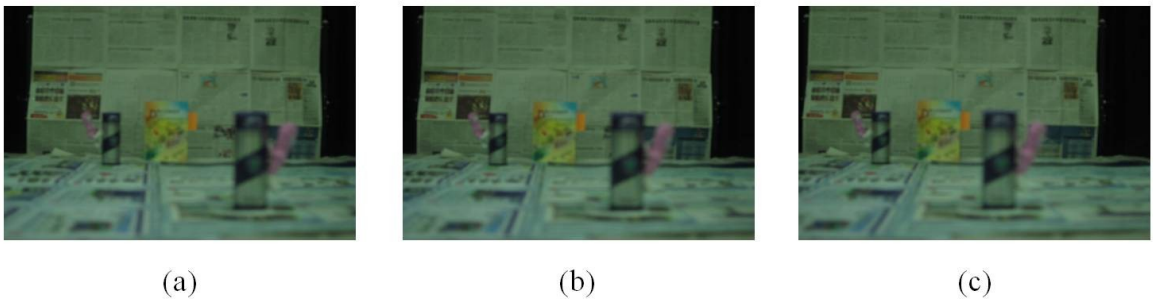


Figure 4-5 Three out-of-focus elemental images (a) left perspective (b) central perspective (c) right perspective



Figure 4-6 Rendered depth map from blurred elemental images

Even for soft matching in DERS, the tolerance is still confined. As a consequence, once the spot of one specific object point is getting bigger, the details that can be treated as the feature will be blended together and hard to be distinguished.

According to the result of simulation and experiment, the importance of a clear and crisp image has been corroborated. Moreover, increasing the f-number to extend the depth of field is not always workable, so the following sections will reveal the results of our HDDR system which elongates the range of depth without increasing the f-number.

4.2 HDDR Depth Map Rendering of Two Depth of Field

In this section, we apply temporal HDDR system to confirm our idea that we can extend the depth of field even using small f-number while capturing. To make the verification simple, we did not use the pitch same as the size of lenslet. Instead, pitch of 1cm is chosen in order to avoid that the disparity might exceed the limitation of DERS.

At first, six elemental images with 2 objects locating at 87cm and 150cm are captured under F/2.8 as shown in Figure 4-7. Two objects are focused individually in two set of elemental images. Therefore, two rendered depth maps in this experiment will include only one well-defined object. Moreover, the stripe floor is utilized to supply plentiful matching candidates. And in the end of chapter 4, the influence of more complicated floor will be testified. As Figure 4-8 and Figure 4-9 shows, the distinct results of two depth maps meet with our postulation that well-outlined contours of objects can be rendered unless they are well focused. Moreover, this effect also corresponds to the concept of high dynamic range (HDR) images. HDR image is adjusting the range of luminance while HDDR system is arranging the range of depth of field. Subsequently, we stack all the parts together.

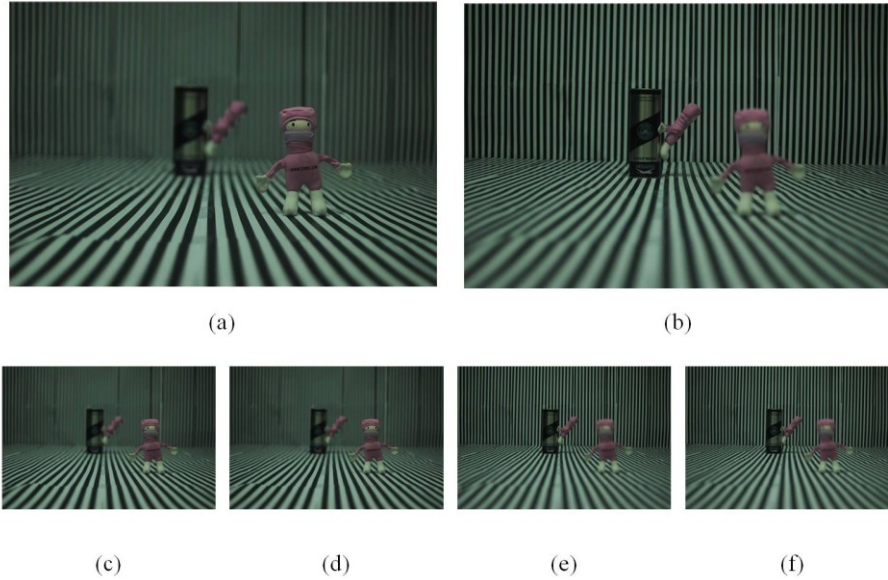


Figure 4-7 Six elemental images of three perspectives and two focal positions and captured under F/2.8 (a)(c)(d) focus at foreground (b)(e)(f) focus at background

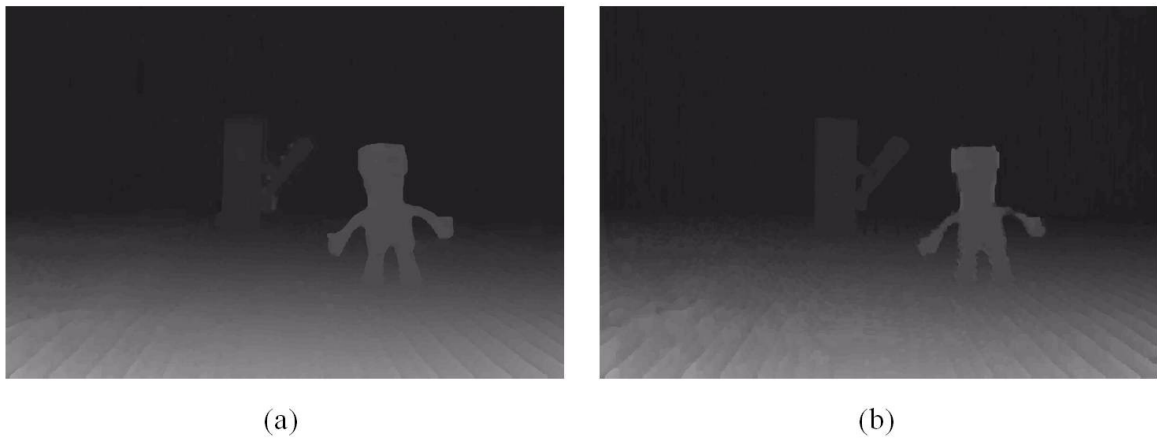


Figure 4-8 Two rendered depth maps (a) focus at foreground (b) focus at background

To fuse these two maps, our target is simply preserving the will-defined objects. Figure 4-10 illustrates process of fusion, in which red and green colors stand for two different depth maps after thresholding. Because we use temporal method, there's no need to do the deviation correction. So we start with edge searching, the corresponding gray levels of two representative focal points are obtained by means of finding the corresponding positions of depth maps, as shown in Figure 4-10 (a)(b). Subsequently, we average the two gray values to determine the threshold and filter the ill-outlined objects, as shown in Figure 4-10 (c).

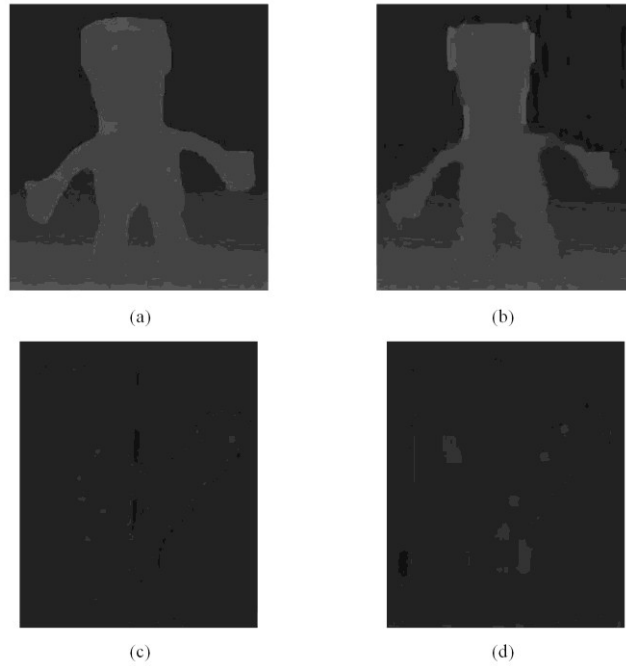
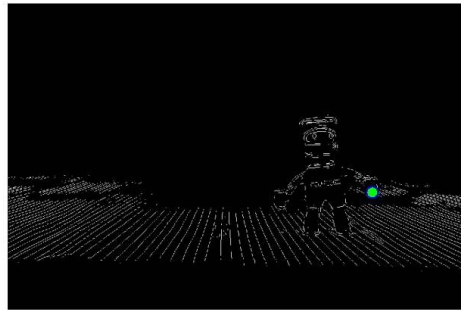
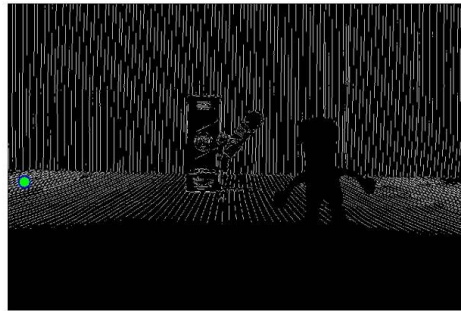


Figure 4-9 Details of objects in rendered depth maps (a)(c) in focus (b)(d) out of focus

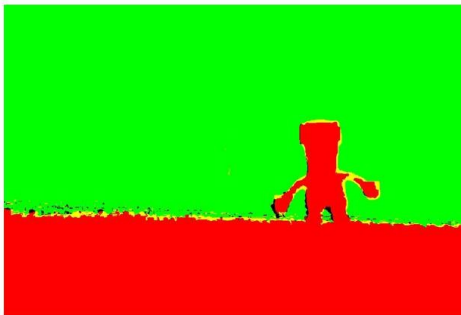
Finally the HDDR depth map can be reconstructed via median filtering, and the result is demonstrated in Figure 4-11 (a). Perceivably, two objects with crisp edge are extracted in HDDR depth map and it is comparable with rendered depth map from elemental images captured by large f-number as shown in Figure 4-11 (b). However, there are some imperfections on the surface of the objects, i.e. darker spots, and these errors might be caused from the noise. Notwithstanding HDDR depth map is generated in a relatively complex method compared to increasing the f-number in this experiment, the capturing time is dramatically reduced by 32 times even for temporal HDDR system. If we use spatial HDDR system, the capturing time can shorter by 64 times than that of larger f-number. In the following section, we use three depth of field to break through the working range of largest f-number ($f/22$) of our camera in experiment.



(a)

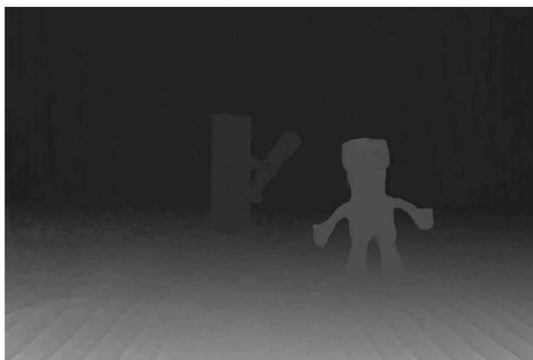


(b)

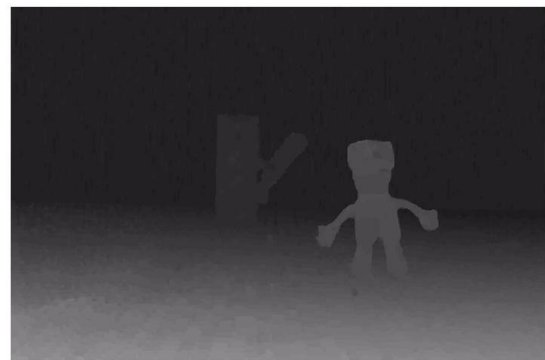


(c)

Figure 4-10 Experiment images during fusion process (a)(b) finding representative focal point (c) two depth maps fusion after thresholding



(a)



(b)

Figure 4-11 Depth maps rendered of (a) HDDR system (b) large f-number (f/22)

4.3 HDDR Depth Map Rendering of Three Depth of Field

In previous section, even though we use 8 times smaller f-number to capturing the elemental images, the rendering of depth map is still not restricted by the shallow depth of field and according to the experimental result, our HDDR depth map is almost identical to the rendered depth with larger f-number. So in this section, we use one additional depth of field to exceed range of large f-number ($F/22$).

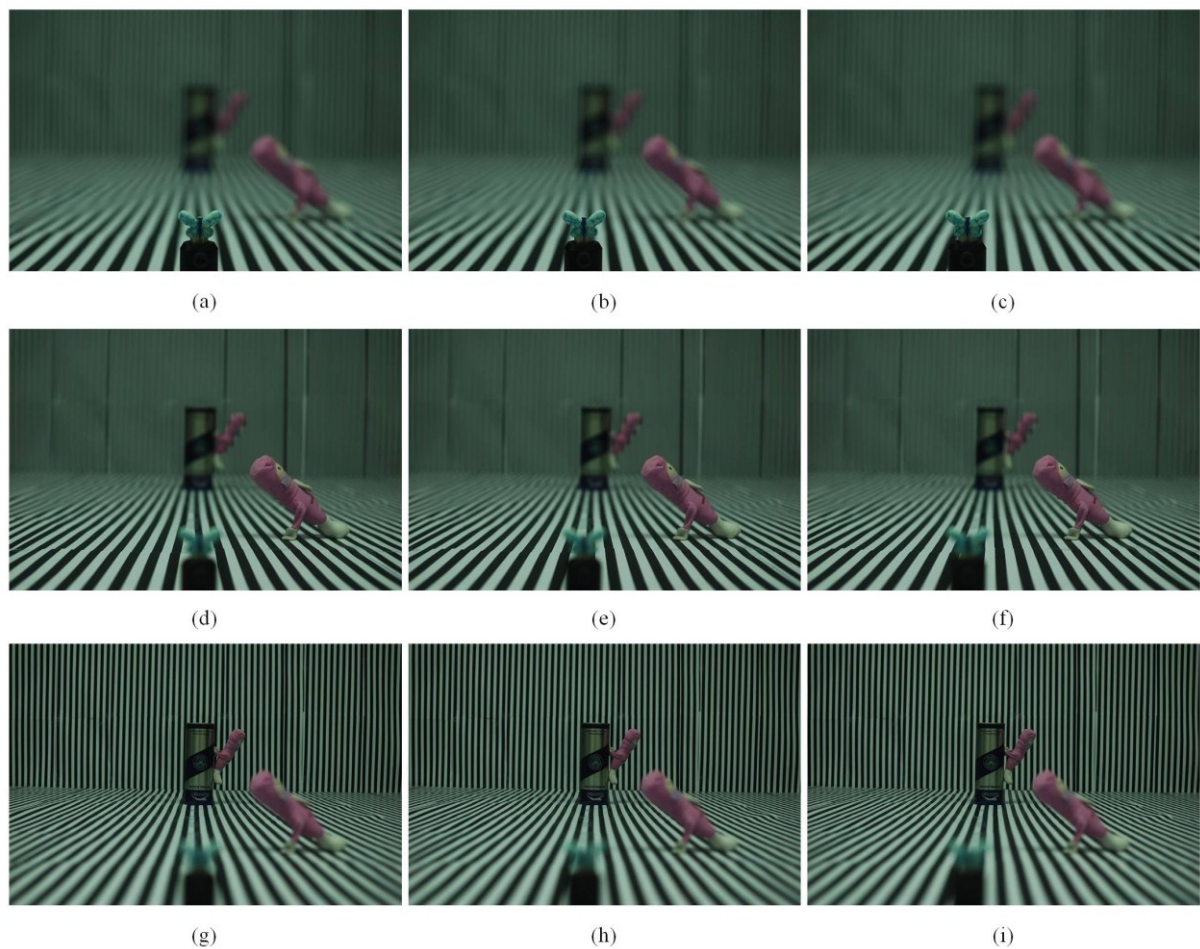


Figure 4-12 Elemental images of three focal positions and captured under $F/2.8$ (a)(b)(c) focus at first object (d)(e)(f) focus at middle object (g)(h)(i) focus at the last object from left, central and right perspective respectively

In this section, we place an additional object, little butterfly, at 35 cm to examine the feasibility in the nearer region and the positions of the other two objects are similar in the

previous experiment. They are set at 76 and 152 cm respectively. First of all, same f-number, $f/2.8$, is applied to capture nine elemental images containing three focal positions as shown in Figure 4-12. Subsequently, every three of them are inputted into DERS to render a depth map. Figure 4-13 illustrates the idea again that the object will be well-defined in the depth map as long as it is focused. According the result in Figure 4-14, we could further verify that blur is one of the factors that govern the accuracy of depth map rendering. From the tendency of the degradation of contours, it is clear that when the object is distant from the depth of field, say the last object while focusing at the first object, the result becomes worse because it blurs more. Likewise, the first object is ill-defined especially when we focus at the last object. Due to the fact that depth of field is a function of object distance, so it shrinks when the objects are placed closer to the camera. This phenomenon particular benefits our HDDR system in near field because f-number has its upper limit for conventional cameras.

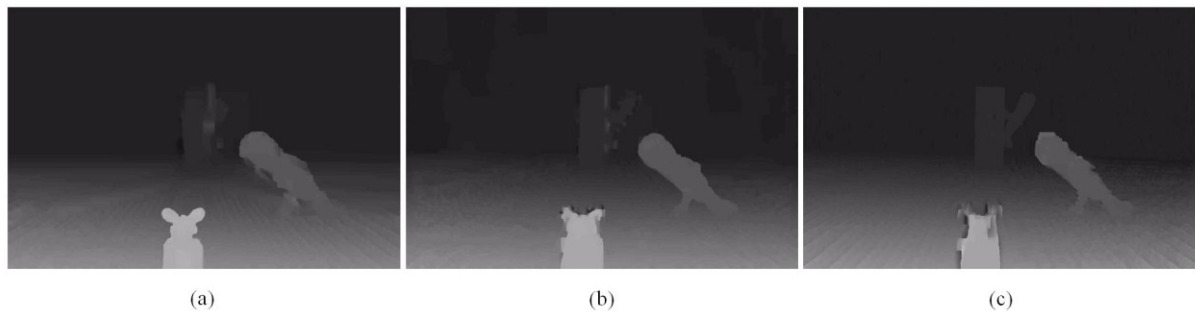


Figure 4-13 Three rendered depth maps (a) focus at first object (b) focus at middle object (c) focus at the last object

By the same steps of fusion elaborated in chapter 3, because we have three depth of field, three representative focal points should be detected as shown in Figure 4-15. Once the corresponding gray levels are found, threshold value can be calculated by averaging two of them. Owing the noise, the thresholding will bring about imperfection combination as illustrated in Figure 4-16. There are many voids lying along the boundaries. Besides, some isolated regions remain after thresholding such as the blue and cyan spots in the magnified image. Unfortunately these redundant spots cannot be eliminated in the following process

because they should have been cut out while thresholding. Hence, when we reconstruct the HDDR depth map, they will leave the darker spots around the first object.

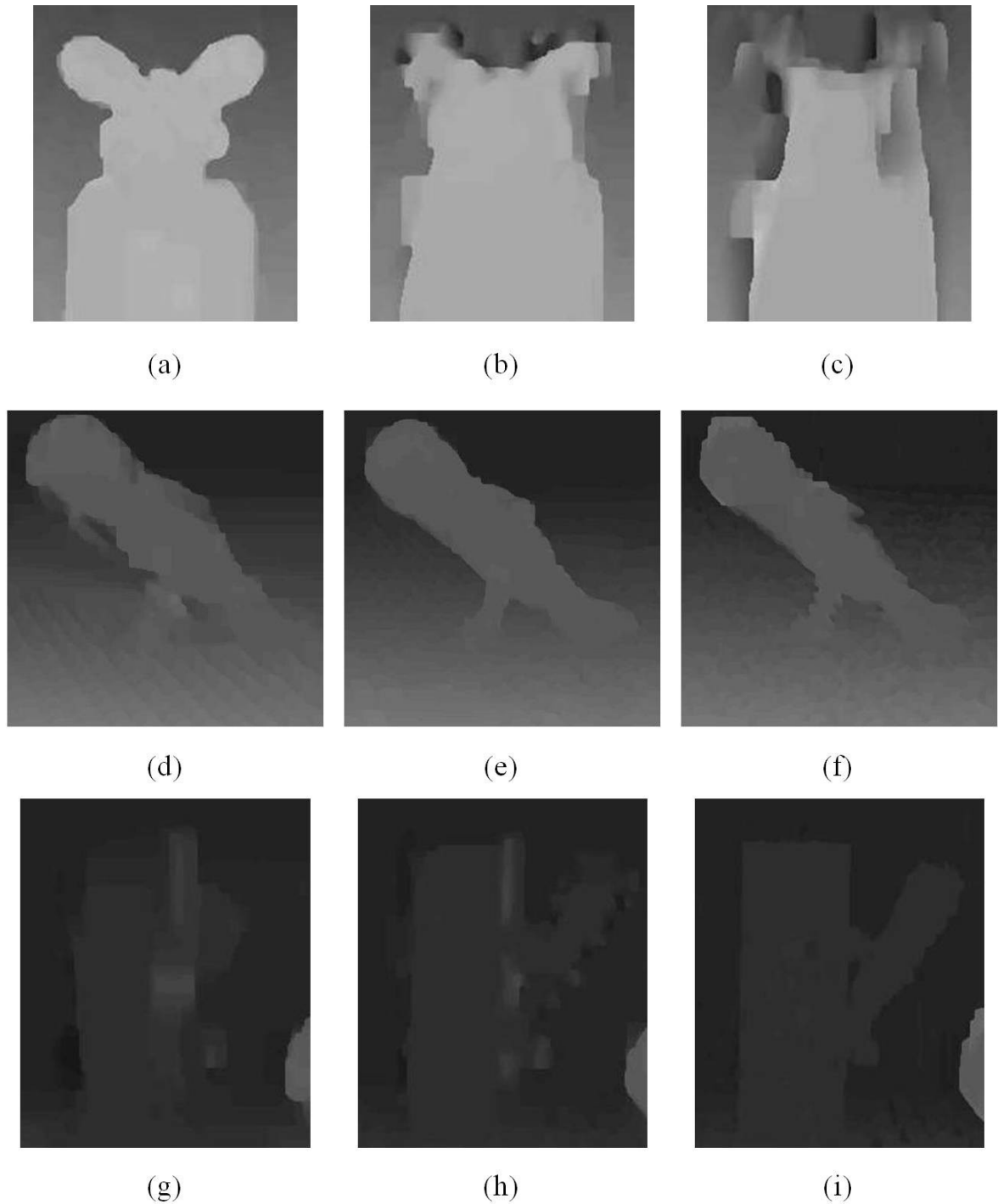
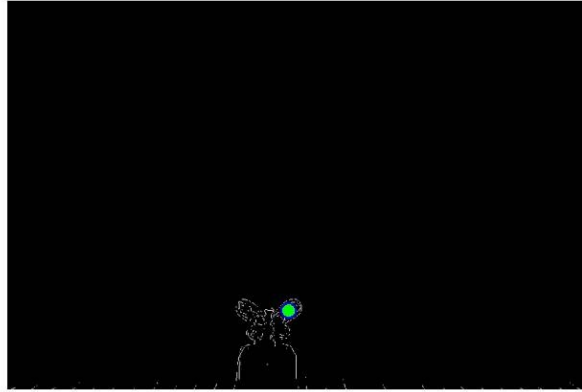
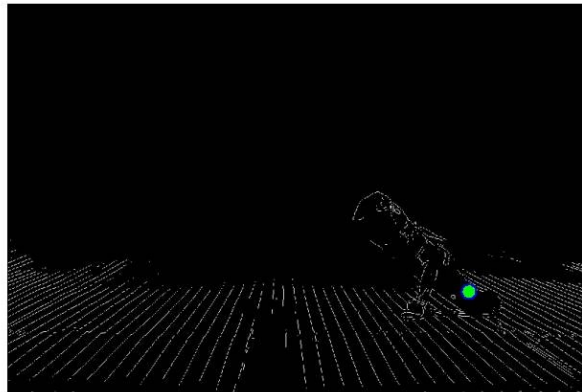


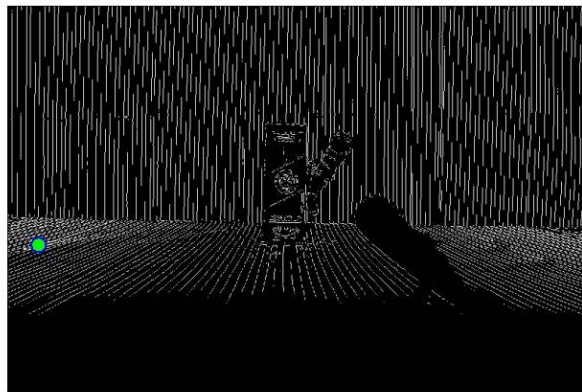
Figure 4-14 Details of objects in rendered depth maps (a)(d)(g) focus at the first object (b)(e)(h) focus at the middle object (c)(f)(i) focus at the last object



(a)



(b)



(c)

Figure 4-15 Experiment images of finding three representative focal points

As shown in Figure 4-17 (a), the surrounding points of the first object in HDDR depth map are worse than that in its origin depth map owing to two reasons. One is described in the previous paragraph. The other is that the lighter regions are the noise of the second depth map. Because the evaluation of edge is judged by the contrast, the object will look ill-outlined if its background is in a mess.

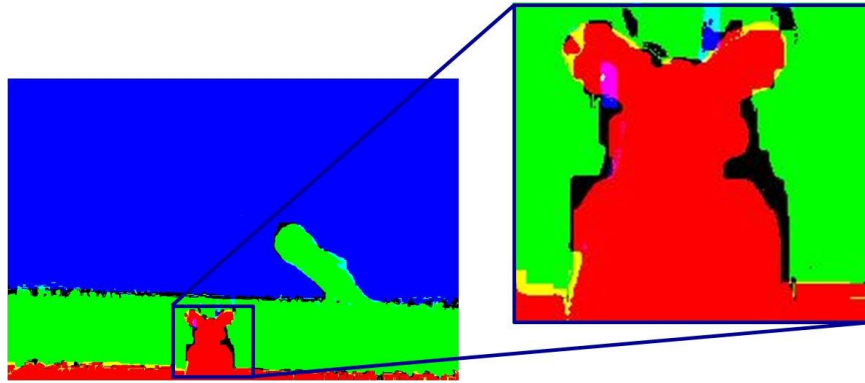


Figure 4-16 Experiment results of fusion and its details

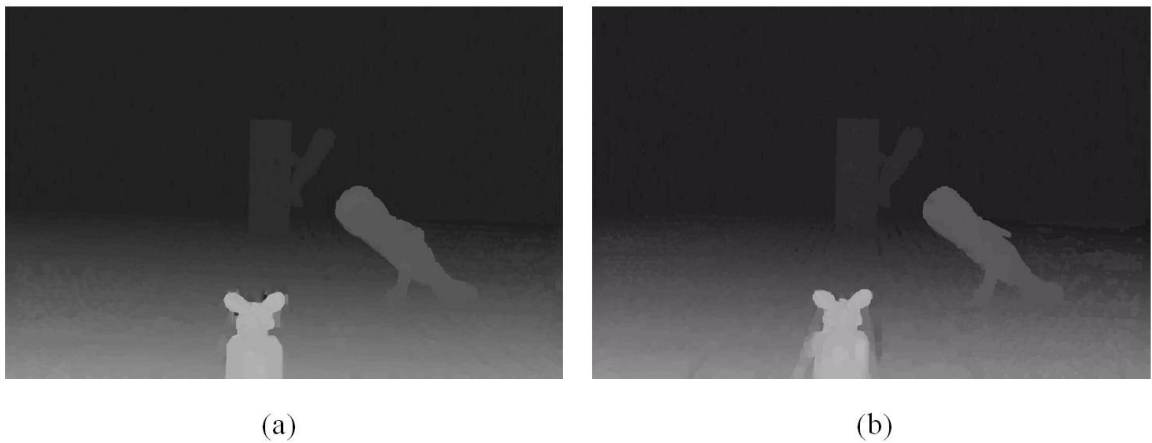


Figure 4-17 Depth maps rendered of (a) HDDR system (b) large f-number ($f/22$)

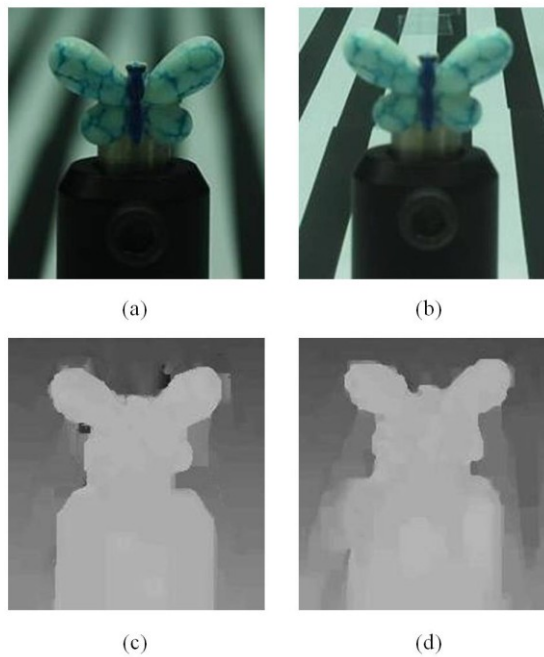


Figure 4-18 Comparison of the first object in color image and depth map of (a)(c) HDDR system (b)(d) large f-number ($f/22$)

Therefore, when we compare the HDDR depth map with the rendered depth map shown in Figure 4-17 (b), the discrepancy of two depth maps is lessened. However, even for the largest f-number of our camera, the scene still cannot be captured all in focus. As shown in Figure 4-18 (b), the veined wings is actually blurred, so the first object in the rendered depth map is slightly fuzzy. And this result can prove that our HDDR system not only surpasses the dynamic range of the capturing with large f-number, but also can be extended to the case of stacking more depth of field so as to render even higher dynamic depth range.

To quantify the working range of different focal design, we utilize Figure 4-4 to judge the range with acceptable degree of blur. We set focal plane at 150 cm and measure the variance of a black-and-white edge. Figure 4-19 shows the concept of point spread function that an ideal point image diverges as it is distant from the focal plane and the smaller f-number, the faster it diverges. Because we hope the error rate of rendered depth map can be less than 10%, the upper bound of variance is around 10 pixels. Accordingly, the working range can be decided as shown in Figure 4-20.

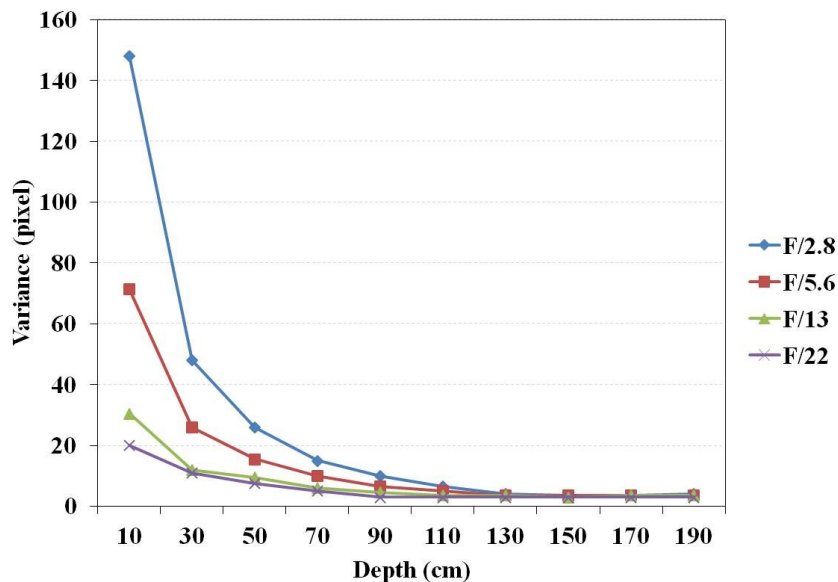


Figure 4-19 Variance of different F/# versus depth

The working range of HDDR system is counted from the first object to the terminal wall (200 cm). Apparently, the working range increase with larger f-number, but working range of

HDDR using small f-number (F/2.8) is even wider than that of the largest f-number (F/22) of our camera. Furthermore, the exposure time is also minimized as illustrated in Figure 4-21. Around 21 times shorter exposure time will benefit the capturing of the instantaneous moments. If spatial HDDR system is implemented, the exposure time will be reduced more and kept the same while stacking more depth of field.

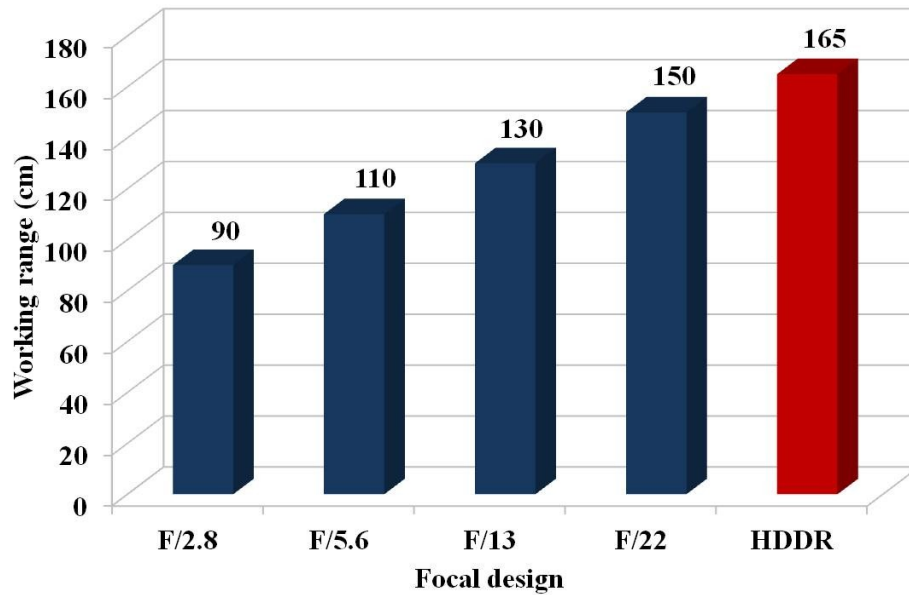


Figure 4-20 Working range of different focal designs

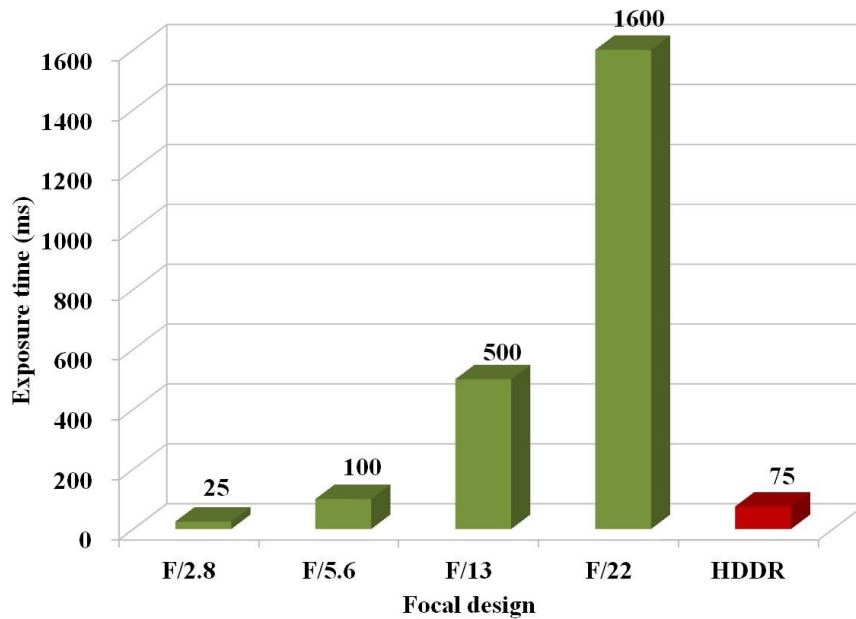


Figure 4-21 Exposure time of different focal designs

To conclude, as long as the render depth map is less vulnerable to noise, the performance of HDDR depth map will be better. However, compared to largest f-number of our camera (F/22), HDDR system not only extends to the wider working range 165 cm (>150 cm), but also minimizes the exposure time by around 21 times shorter. And in the following section, we will change the pitch to a reasonable quantity, the size of lens, to verify the feasibility of applying lens array. Moreover, undersigned background is utilized in the scene to meet with more general situation in real word.

4.4 From Temporal to Spatial HDDR System

Although we've proved the possibility to accumulate several depth of field to render a HDDR depth map which surpasses the range by capturing with large f-number, there are still some factors without carefully examined Hence, we'll change three things in this section: complex floor and terminal wall, larger pitch, and the arrangement of objects in the depth of field.

Black and white striped patterns contain regular and strong features, but in real world it is an impractical background. As a result, we use posters and newspapers in order to increase the complexity of the floor and terminal wall.

Because we use moving camera to simulate the result of capturing by lens array, the pitch of moving should be equal to the size of lenslet at least. Therefore, we increase the pitch from 1cm to 5cm. In fact, the movement of 5 cm might be too large for near objects and it would lead to more matching errors, so the objects in this experiment are placed farther.

Thirdly, objects locating in the same depth of field will be regarded as no depth difference for sweep focus system due to the same point spread function, but as far as stereo matching algorithm is concerned, there still exists different disparity between these objects. Accordingly, we put two objects in the same depth field to verify this superiority.

Finally, to keep the system straightforward, two depth of field are utilized in this experiment.

4.4.1 Temporal HDDR System

As shown in Figure 4-22, the elemental images are captured in the same height because we use temporal HDDR system, and the first object and the last two objects are placed in the different depth of field. Besides, the floor and the terminal wall are full of irregular patterns. Consequently, the variation of rendered depth map is larger than we use striped pattern, as shown in Figure 4-23 (a)(b). It is obvious that some regions are totally wrong such as the very front in Figure 4-23 (a). However, some of the regions will be filtered after thresholding because these gray levels do not belong to their segments. When they are purged, the voids will be reconstructed by median filter. As a result, Figure 4-23 (c) demonstrates the HDDR depth map that reserves the well-defined objects and “repairs” some of the mismatching regions. But the error of first object in Figure 4-23 (b) is not well eliminated, the quality of HDDR depth is degraded accordingly. Compared to the out-of-focus regions in rendered depth maps, the contour in HDDR has been improved a lot.

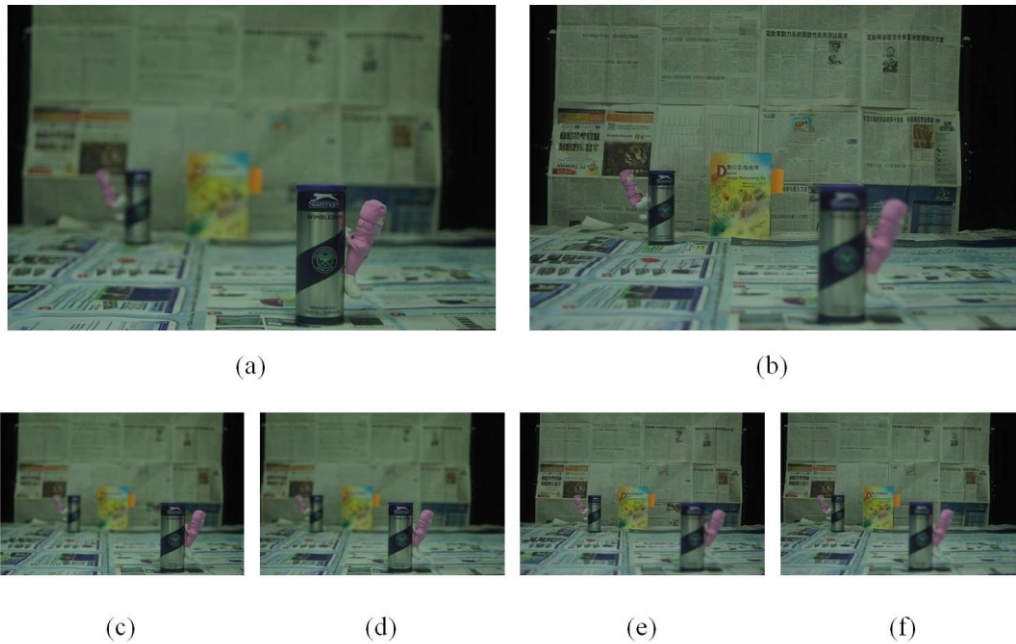


Figure 4-22 Six elemental images of three perspectives and two focal positions and captured under F/2.0 (a)(c)(d) focus at foreground (b)(e)(f) focus at background



(a)



(b)



(c)

Figure 4-23 Two rendered depth map and the result after fusion (a) focus at foreground (b) focus at background (c) HDDR depth map

4.4.2 Spatial HDDR System

In terms of temporal HDDR system, the capturing time is proportional to the number of depth of field. However, the capturing time is independent of the number of depth of field for spatial HDDR system. Therefore, elemental images should be arranged on the sensor plane instead of being arranged along the time. As Figure 4-24 (a) (b) illustrate, both of the focal plane and perspective are altered. So when we combine the two rendered depth maps, the deviation in height should be correct at the very beginning. In addition, similar problem of mismatching appears in two rendered depth maps as shown in Figure 4-25 (a)(b). If these regions cannot be removed, it will cause a disagreeable spot in HDDR depth map such as the right side of Figure 4-25 (c).

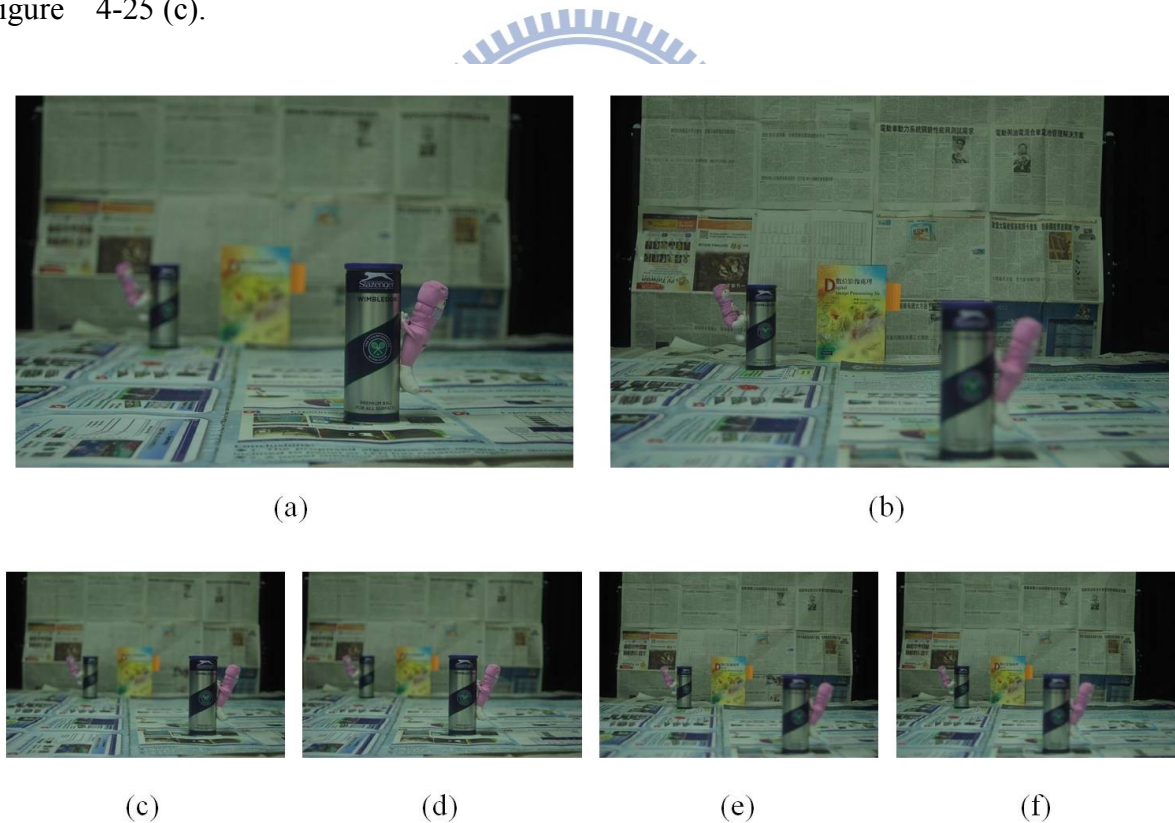
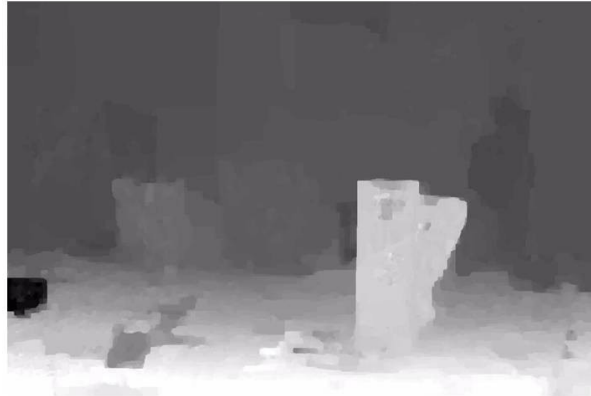


Figure 4-24 Six elemental images of three perspectives and two focal positions and captured under F/2.0 (a)(c)(d) focus at foreground (b)(e)(f) focus at background



(a)



(b)



(c)

Figure 4-25 Two rendered depth map and the result after fusion (a) focus at foreground (b) focus at background (c) HDDR depth map

However, the contour of first object might vary with the different perspective, so there might be sort of distortion while shifting in deviation correction. As the pitch of lenslet becomes smaller, the difference of perspectives will be mitigated. Furthermore, the scene is captured by single shot in spatial HDDR system, so it has the potentiality to deal with moving objects with

very short exposure time. This capability is impossible for so called “all-in-focus” images by simply increasing the f-number. Lastly, Table 4-1 summarizes the uniqueness of two HDDR systems. It should be note that according to the experimental results in this section, both of the two HDDR systems are able to distinguish different objects in the same depth of field with respect to depth while sweep focus system need depth-varying point spread functions.

	Temporal HDDR system	Spatial HDDR system
# of shots	proportional to # of depth of field	one
resolution sacrifice	decrease one third	decrease depending on # of depth of field
complexity of algorithm	simpler	more complicate
Issues	fail in moving objects	distortion from different perspectives

Table 4-1 Comparison of two HDDR systems

4.5 Discussion

With regard to HDDR system, there are two major issues. One is the limited matching range in near field. When the object distance shrinks, the depth of field becomes shorter as well. Accordingly, we need more lenslets to extend the depth. However, because the field of view is limited, the range for stereo matching is therefore confined. In other words, we cannot ensure the whole scene is totally captured by every lenslet we add. In Figure 4-26, matching range stands for the region captured by at least three lenslet while blind range means the region(s) that would never be captured. So if we change the number of elemental images for stereo matching, maybe we can extend the matching range. But it is impossible to reduce the blind range unless we increase the field of view. Nevertheless, to increase the field of view will bring about severe lens aberration. Even though we do not increase the field of view, the lens aberration may still influence our elemental images because the objects deviate from the optical axis of some lenslets. The other issue is the reliable stereo matching algorithm. For

disparity-based system, stereo matching is always needed to improve. Even though our HDDR system has dealt with the matching error in blur region, the issue of different perspectives still challenges the yield rate because the same feature point would not appear in the two or three elemental images simultaneously. Moreover, occlusion also happens with different perspective as shown in Figure 4-27 [55]. Therefore we cannot make sure that the render depth map is correct even if the elemental images are all in focus. Once the quality of conventional depth maps are enhanced, HDDR depth can be enhanced as well because there will be less misjudgments while thresholding.

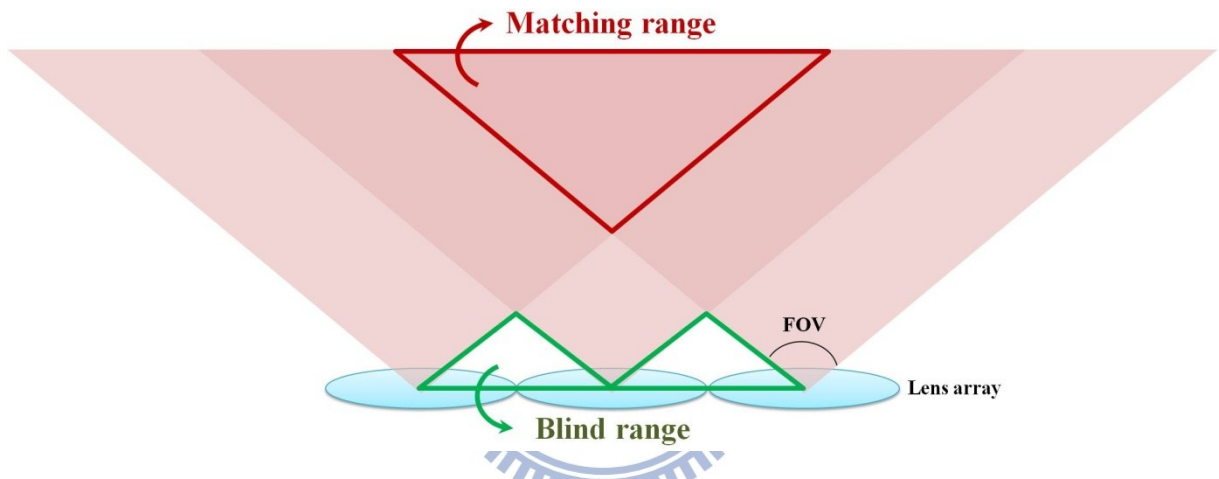


Figure 4-26 Matching range and blind range in near field

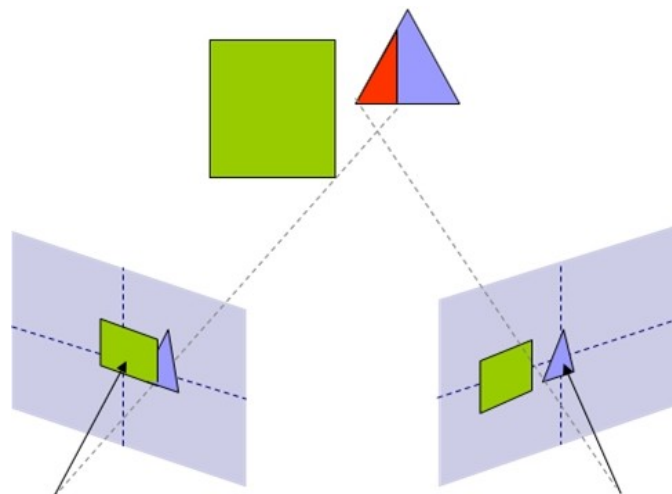


Figure 4-27 Occlusion geometry

Chapter 5

Summary

5.1 Conclusion

3D content plays an important role in 3D technology, so how to capture the depth information is a target that many researchers are digging into. According to the prior arts in chapter one, single camera system based on analysis of disparity is the most practical system, but it requires all-in-focus elemental images to ensure the feature matching. However, extending the depth of field by increasing the f-number is not always suitable. Especially for dimming environment and the image content with movement, capturing with larger f-number confronts the dilemma of exposure time. If the exposure time is insufficient, the image information will be lost. On the other hand, if the exposure time is too long, the ghost image will bring about the blurred images. Two situations cause the imperfections in the rendered depth map because of mismatch in stereo matching algorithm. Hence, we extend the working range by stacking each depth of field instead of increasing the f-number. This concept originates from the high dynamic range (HDR) images which use limited contrast ratio to represent higher luminance difference. Accordingly, our system is named after a similar term of high dynamic depth range (HDDR). The idea of HDDR system can be fulfilled in two manners: temporal or spatial multiplex.

Regarding the algorithm, if DERS can be less vulnerable to noise, the performance will be enhanced. Besides, the limitation of DFEET stems from the texture and shape of objects. Non-textured regions will contribute no edge information, which will greatly influence the threshold value and cannot filter out the ill-defined objects. As for the shape of object, void with the small aspect ratio region is hard to be reconstructed. In general, current stereo matching algorithm cannot render a depth map with all the details as the color image, so our

fusion process is still able to generate the HDDR depth map with acceptable quality.

In chapter four, we've demonstrated the experimental results of both of the systems, and from the HDDR depth maps, it's been proved that the depth range of HDDR system (165 cm) goes beyond that of the result rendered by largest f-number of our camera in the experiment (150 cm) and the capturing time can be minimized by at least 21 times. Furthermore, the feasibility of lens array has been examined by using the moving pitch same as the size of the lenslet. And the distinction of two systems is that temporal HDDR system maintains the resolution of elemental images as the number of depth of field increases, while spatial HDDR system is capable of capturing moving objects with single shot and very short exposure time.

In the end, our HDDR system is compared with the prior arts and the result is illustrated in Table 5-1. The first compared aspect is the size of system which is quite important for commercialization. In addition, image categories should not be restricted and the depth rendering by optical and geometrical measurement is more favorable and reliable. And the last two compared factors are working range and capturing time. The image quality of both color images and depth maps are improved by pursuing higher dynamic range, so in depth maps, the dynamic range, that is working range, is represented as the depth extension. As for capturing time, it goes without saying that the shorter it is, the better the system will be.

Actually, the working range is restricted by the accuracy of stereo matching algorithm and the sensor resolution, no matter which type of the HDDR system is. The disparity increases as the density of resolution increases, so when applying stereo matching algorithm, the risk of mismatch also increases. Therefore, the realization of wonderful depth map rendering requires more robust techniques of feature extraction and matching. Optically, we'll propose an idea to well utilize the tolerant range of disparity in following part, future work.

System		Size of system	Diversity of image category	Reliance of depth values	Working range	Capturing time
Computer-based 2D-3D conversion		☹	☹	☹	☹	☹
Multi-camera system		☹	☹	☹	☹	☹
Single-camera system	Integral image	☹	☹	☹	☹	☹
	Plenoptic camera	☹	☹	☹	☹	☹
	Sweep focus	☹	☹	☹	☹	☹
HDDR		☹	☹	☹	☹	☹




 satisfactory
 acceptable
 unsatisfactory

Table 5-1 Comparison table of our HDDR system with the prior arts

5.2 Future Work

In future work, following the concept of HDDR system, we can apply it in very near field. To reduce the disparity, we have to minimize the size of lenslet. Therefore, liquid crystal lens array is a candidate. Moreover, because high dynamic range image also includes an idea of delicate resolution of luminance, we are eager to render a depth map with not only large depth range but also fine depth resolution. Once this kind of depth map can be generated, we can resize depth range and zoom in the scene.

5.2.1 Liquid Crystal Lens Array

Liquid crystal (LC) is a birefringence material whose refractive index varies as the polarization, and it can be controlled by electrical field. Therefore, the character is suitable for phase modulator. LC lens is a kind of GRIN lens and its lens function can be switched in and off as illustrated in Figure 5-1 [56].

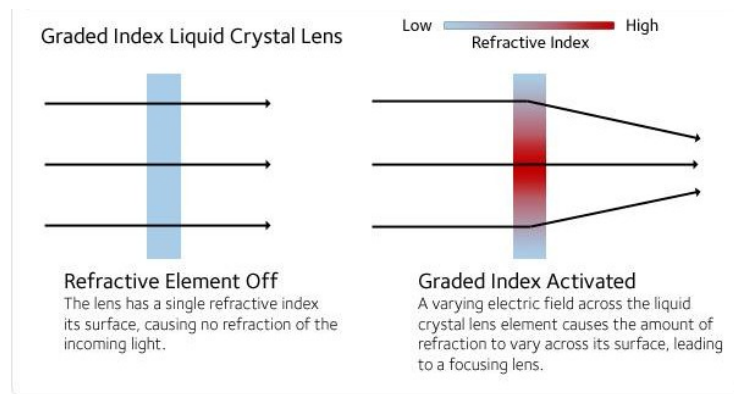


Figure 5-1 Mechanism of GRIN lens

By adjusting the applied voltage, the distribution of effective refractive index also can be changed so as to alter the focal length. Besides, LC lens is very small and thin due to its high birefringence. Hence, replacing conventional lens with LC lens benefits our HDDR system in three aspects. First of all, the arrangement of the depth of field can be optimized via the tunable focus capability of LC lens. Secondly, if customers want to shoot conventional 2D images, they don't have to take off the lens array but shut down the lens function electrically. Thirdly, smaller pitch of lens array can reduce the disparity of objects. In other words, we are capable of capturing nearer objects in the scene. Furthermore, many small operations are carried out by endoscopes, so if tiny LC lens array with our HDDR system can be embedded with endoscopes, maybe we can use HDDR depth map to build a 3D model of the tissue or the tumor as shown in Figure 5-2 [57]. The 3D model would help surgeons operate the surgery and leave smaller wounds.

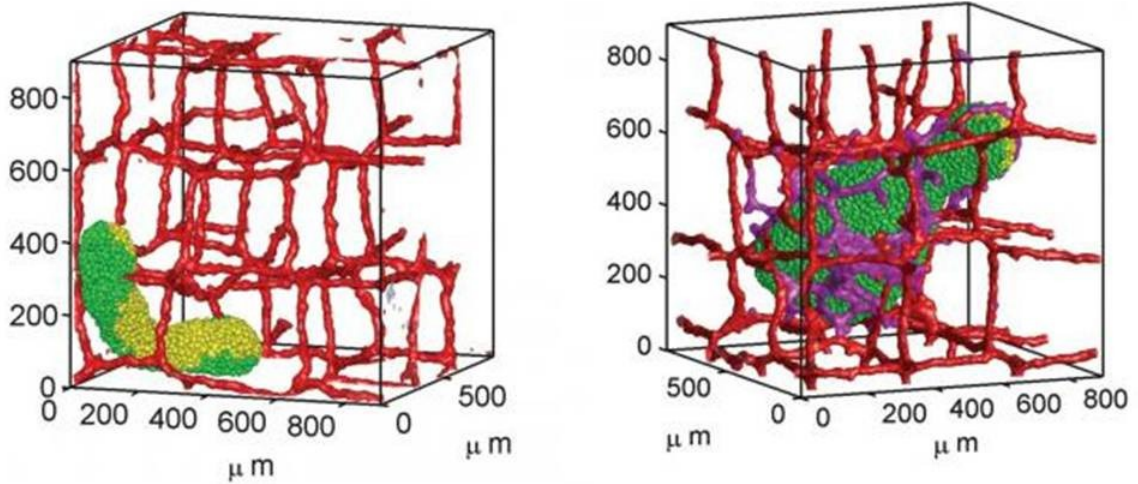


Figure 5-2 Scheme of the 3D model of a tumor growing

5.2.2 Fine Depth Resolution

Because of the limitation of the searching range in stereo match algorithm as well as the size of pixel, the depth rendering is confined within a specific range of disparity. DERS can tolerate the disparity within 100 pixels, so the larger the pitch of lens array is, the farther positions can be rendered correctly. On the other hand, if the objects is remote from the camera, the disparity smaller than one pixel will not be recorded. As a result, given the size of pixel and the pitch of lens array, the bounded depth range can be calculated as the following equations.

$$\text{Disparity } (\delta) = pg/d \quad (39)$$

where p and g are the pitch and the gap between lens array and sensor respectively, and d is depth from lens array. Because the disparity limit is from 1 pixel to 100 pixels, so the depth range becomes

$$\text{Depth range (DR)} = 100pg - pg = 99pg \propto p \quad (40)$$

Accordingly, we can design a system to generate a depth map with fine depth resolution by adjusting the pitch as shown in Figure 5-3 and Figure 5-4. We can use three columns of elemental images (say column3 to column5) to render one HDDR depth map and change

another set of three columns of elemental images (say column2, column4, and column6) to render another HDDR depth map. For example, if HDDR-2 region in Figure 5-3 is the conventional range of one depth map with 8-bit gray levels, every object in HDDR-3 region would be regarded as no depth different because they don't have disparity. However, if we can zoom in the scene along the depth, the objects in HDDR-3 region is no more at the same depth. Accordingly, we can use this technique to generate "high definition" depth maps and resize the depth as conventional 2D images.

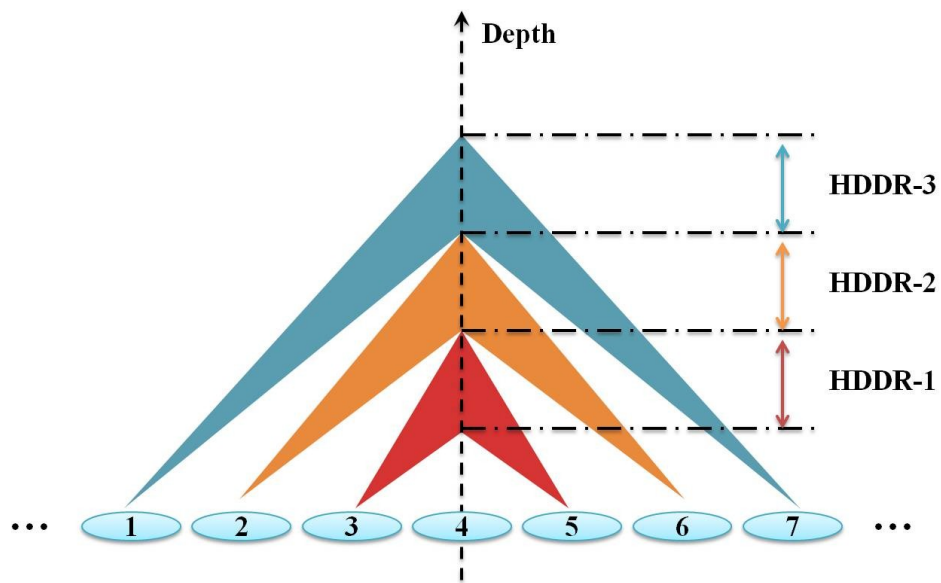


Figure 5-3 Horizontal scheme of fine depth resolution design

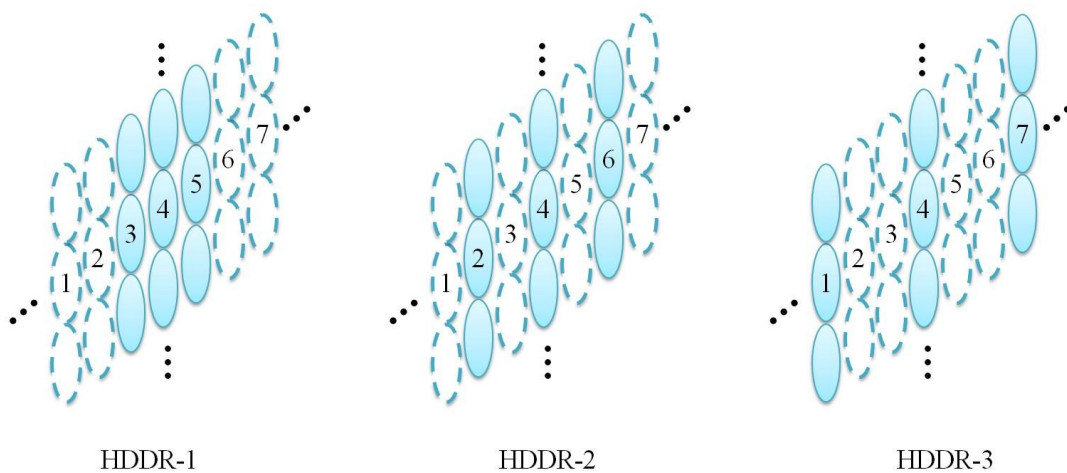


Figure 5-4 Distribution of lens array with different rendering depth ranges

Reference

- [1] http://en.wikipedia.org/wiki/Digital_camer
- [2] http://en.wikipedia.org/wiki/Pulfrich_effect
- [3] H. Murata, et al., "A Real-Time 2-D to 3-D Image Conversion Technique Using Computed Image Depth", *SID SYM*, Vol. 29, pp. 919-922, 1998.
- [4] P. Harman, et al., "Rapid 2D to 3D Conversion", *In Proc. SPIE, Stereoscopic Displays and Virtual Reality Systems IX*, Vol. 4660, pp. 78-86, 2002.
- [5] S. Battiato, et al., "Depth-Map Generation by Image Classification", *In Proc. SPIE, Three-Dimensional Image Capture and Applications VI*, Vol. 5302, pp. 95-104, 2004.
- [6] J. Li, et al., "A Novel Image-Based Rendering System with a Longitudinal Aligned Camera array", *EUROGRAPHICS*, 2000.
- [7] R. Furukawa, et al., "One-shot entire shape acquisition method using multiple projectors and cameras", *In Fourth (PSIVT) of IEEE Computer Society*, pp. 107-114, 2010.
- [8] R. Furukawa, et al., "Multiview Projectors/Cameras System for 3D Reconstruction of Dynamic Scenes", *In 4DMOD-Workshop on Dynamic Shape Capture and Analysis 2011*, hal-00675085, version 1, 2012.
- [9] C. Zhou, et al., "Focal Sweep Camera for Space-Time Refocusing", Columbia University Computer Science Tech. Reports, 2012.
- [10] O. Cossairt and S. Nayar, "Spectral Focal Sweep: Extended Depth of Field from Chromatic Aberrations", *IEEE Conference on Computational Photography*, pp. 1-8.
- [11] http://en.wikipedia.org/wiki/Plenoptic_illumination_function
- [12] E. H. Adelson and J. Y. A. Wang, "Single Lens Stereo with a Plenoptic Camera", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 99-106, 1992.
- [13] R. Ng, et al., "Light Field Photography with a Hand-held Plenoptic Camera", Stanford Tech Report, CTSR 2005-02, 2005.
- [14] A. Stern and B. Javidi, "Three-Dimensional Image Sensing, Visualization, and Processing Using Integral Imaging", *In Proc. IEEE*, Vol. 94, No. 3, 2006.
- [15] J. Arai, et al., "Gradient-index lens-array method based on real-time integral photography for three-dimensional images", *Applied Optics*, Vol. 37, No. 11, 1998.
- [16] A. Stern and B. Javidi, "Three-dimensional imaging sensing, visualization, and processing using integral imaging", *In Proc. IEEE on 3-D technologies for imaging and display*, 2006.
- [17] Y. T. Lim, et al., "Analysis on enhanced depth of field for integral imaging microscope", *Optics Express*, Vol. 20, No. 21, 2012.
- [18] J. S. Jang and B. Javidi, "Large depth-of-focus time-multiplexed three-dimensional

- integral imaging by use of lenslets with nonuniform focal lengths and aperture sizes”, *Optics Letters*, Vol. 28, No. 20, 2003.
- [19] M. C. Rau¹, et al., “Progress in 3-D Multiperspective Display by Integral Imaging”, *In Proc. IEEE*, Vol. 97, No. 6, 2009.
- [20] <http://www.wisegeek.com/what-is-the-difference-between-monocular-and-binocular-vision.htm>
- [21] N. Holliman, “3D Display System”, 2005.
- [22] L. Hill, et al., “Invited paper: 3-D Liquid Crystal Displays and Their Applications”, *IEEE*, Vol. 94, No. 3, 2006.
- [23] J. Mansson, “Stereovision: A model of human stereopsis”, Lund University Cognitive Science, Tech. Report., 1998.
- [24] E. Edirisinghe, et al., “Stereo image, an emerging technology”, *SSGRR*, L’Aquila, Italy, 2000.
- [25] <http://www.vision3d.com/stereo.html>
- [26] G. Woodgate, et al., “Flat panel autostereoscopic displays-characterization and enhancement”, *In Proc. SPIE*, Stereoscopic Displays and Virtual Reality Systems VII, Vol. 3957, pp. 153-164, 2000.
- [27] H. Morishima, et al., “Rear cross lenticular 3D display without eyeglasses”, *In Proc. SPIE*, Stereoscopic Displays and Virtual Reality Systems VII, Vol. 3295, pp. 193-202, Apr. 1998.
- [28] Y. Schechner and N. Kiryati, “Depth from Defocus vs. Stereo: How Different Really Are They?”, *IEEE International Journal of Computer Vision*, Vol. 39, No. 2, pp. 141-162, 2000.
- [29] C. Zhou, et al., “Coded Aperture Pairs for Depth from Defocus”, *IEEE International Conference on Computer Vision*, 2009.
- [30] J. Hong, et al., “Three-dimensional display technologies of recent interest: principles, status, and issues. [Invited]”, *Applied Optics*, Vol. 50, No. 34, 2011.
- [31] <http://ucalgary.ca/pip369/mod4/depthperception/oculomotor>
- [32] T. C. Shen, “Autostereoscopic 2D-3D Switching Display with Multi-Electrically Driven Cylindrical Liquid Crystal Lens”, Master, Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan, ROC, 2009.
- [33] <http://ucalgary.ca/pip369/mod4/depthperception/monocular/pictorial>
- [34] <http://ucalgary.ca/pip369/mod4/depthperception/movement>
- [35] <http://en.wikipedia.org/wiki/Microscopy>
- [36] W. J. Smith, “Modern Optical Engineering”, Fourth edition, McGraw-Hill, New York, 2008.
- [37] C.M. Sparrow, “On Spectroscopic Resolving Power”, *Astrophysical Journal*, Vol. 44, pp.76-86, 1916.

- [38] E. Hecht, "Optics", Fourth edition, Addison Wesley, San Francisco, 2002.
- [39] C. A. Bennett, "Principle of PHYSICAL OPTICS", 吳忠義等譯，滄海書局，台中，民國九十八年。
- [40] http://en.wikipedia.org/wiki/Hyperfocal_distance
- [41] http://en.wikipedia.org/wiki/Compound_microscope#Compound_microscope
- [42] http://en.wikipedia.org/wiki/Dynamic_range
- [43] F. Houlmann and S. Metz, "High Dynamic Range Rendering in OpenGL", UTBM.
- [44] http://en.wikipedia.org/wiki/Dodging_and_burning
- [45] J. Kuang, et al., "Evaluating HDR Rendering Algorithm", Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, New York, 2006.
- [46] K. Devlin, "A review of tone reproduction techniques", Tech. Report CSTR-02-005, Department of Computer Science, University of Bristol, 2002.
- [47] <http://en.wikipedia.org/wiki/HDR>
- [48] J. W. Goodman, "Fourier Optics", Third edition, ROBERTS & COMPANY, Colorado, 2005.
- [49] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV", *In Proc. SPIE, Stereoscopic Displays and Virtual Reality System XI*, 2004.
- [50] <http://mpeg.chiariglione.org/about>
- [51] O. Stankiewicz, et al., "A soft segmentation matching in Depth Estimation Reference Software (DERS) 5.0", MPEG, M17049, 2009.
- [52] A. Olofsson, "Modern Stereo Correspondence Algorithms: Investigation and evaluation", Master, Institute of systemteknik, Linköping University, Linköping, Sweden, 2010.
- [53] Gonzales and Woods, "Digital image processing, 3/e", 繆紹綱譯，台灣培生教育出版股份有限公司，台北，民國九十八年。
- [54] S. Lösch, "Depth from Blur-Combining Image Deblurring and Depth Estimation", Master, Department of computer science, Saarland University, Saarbrücken, Germany, 2009.
- [55] http://www.cse.yorku.ca/~sizints/ctf_stereo/ctf_stereo_main.html
- [56] <http://www.vdwoxford.org/ourwork/#lenstech>
- [57] <http://biomedicalcomputationreview.org/content/3d-angiogenesis-modeled>