

EXTENDING LIGHT FIELD CAMERA CAPABILITIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF
ENGINEERING AND NATURAL SCIENCES
OF ISTANBUL MEDIPOL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
ELECTRICAL, ELECTRONICS ENGINEERING AND CYBER SYSTEMS

By
Muhammad Umair Mukati
August, 2017

EXTENDING LIGHT FIELD CAMERA CAPABILITIES

By Muhammad Umair Mukati

August, 2017

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Prof. Dr. Bahadır Kürşat Güntürk(Advisor)



Assis. Prof. Dr. Muhammed Fatih Toy



Assoc. Prof. Dr. Taner Eski

Approved by the Graduate School of Engineering and Natural Sciences:



Prof. Dr. Talip Alp

Director of the Graduate School of Engineering and Natural Sciences

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MUHAMMAD UMAIR MUKATI

Signature : 

ABSTRACT

EXTENDING LIGHT FIELD CAMERA CAPABILITIES

Muhammad Umair Mukati

M.S. in Electrical, Electronics Engineering and Cyber Systems

Advisor: Prof. Dr. Bahadır Kürşat Güntürk

August, 2017

A traditional camera captures an image by projecting a scene on a two-dimensional image sensor plane regardless of the direction light rays coming onto the sensor. A light field camera, in contrast, records light rays in different directions separately. This allows post-capture control of the imaging parameters, such as focal distance and aperture size. With a light field camera, it is possible to focus at different depths, change depth of focus, change the perspective, and estimate the depth, all computationally after recording the light field in a single shot.

There are two major limitations of a light field camera: small aperture size and low spatial resolution. In this thesis, various approaches to address these limitations are presented. Specifically, a multi-capture method to extend the size of the aperture, a hybrid system that includes a regular and a light field sensor to address the spatial resolution issue, and a micro-scanning super-resolution method to improve the spatial resolution are presented.

Keywords: Light field camera, resolution enhancement, synthetic aperture, computational photography.

ÖZET

IŞIK ALAN KAMERALARININ KABİLİYETLERİNİN ARTTIRILMASI

Muhammad Umair Mukati

Elektrik-Elektronik Mühendisliği ve Siber Sistemler, Yüksek Lisans

Tez Danışmanı: Prof. Dr. Bahadır Kürşat Güntürk

Ağustos, 2017

Geleneksel bir kamera, ışınların sensöre geliş yönünü gözönüne almadan, sahneyi iki boyutlu bir görüntü düzlemine aktarır kaydeder. Diğer taraftan, ışık alan kamerası, ışınların farklı yönlerden gelen miktarlarını ayrı ayrı kaydeder. Bu da, odak mesafesi ve diyafram açıklığı gibi görüntüleme parametrelerin çekim sonrası kontrolüne imkan tanır. Işık alan kamerası ile, farklı derinliklere odaklama, odak genişliğini değiştirme, bakış açısını değiştirme, derinliği kestirme gibi işlemleri, tek bir çekim sonrası hesaplamalı olarak yapmak mümkündür.

Işık alan kameralarının iki temel kısıtlaması vardır: küçük diyafram açıklığı ve düşük uzamsal çözünürlük. Bu tezde, bu kısıtlamaları gidermeye yönelik çeşitli yaklaşımlar sunulmaktadır. Özel olarak, diyafram açıklığını arttıran çok çekimli bir metod, uzamsal çözünürlüğü arttıran hibrid (geleneksel ve ışık alan sensöründen oluşan) bir görüntüleme sistemi ve uzamsal çözünürlüğü arttıran mikro-taramalı süper-çözünürlük metodu sunulmuştur.

Anahtar sözcükler: Işık alan kamerası, çözünürlük arttırma, sentetik diyafram açıklığı, hesaplamalı görüntüleme.

Acknowledgement

First of all I want to thank Almighty Allah who has made me able to complete my research work.

This work would not have been possible without the patience, support and knowledge offered to me by my supervisor Prof. Bahadır Kürşat Güntürk. Through his constant supervision, I have been able to build up my research work to a stage that I am able to write this thesis. Throughout my first year in this program, he helped me in laying the foundation of my concepts in this area with his constant guidance. It was only possible due to his supervision, that with the end of my MS program I have been able to publish some high impact research work.

Finally, a special thanks to my family and friends for their constant support specially to my lab-partner Zeshan Alam, who always spared time for a fruitful research discussion which helped me to reach to a solution to some of my research problems.

The work in this thesis is supported by the TUBITAK Project 114E095.

Contents

1	Introduction	1
1.1	Traditional cameras	2
1.2	Light field cameras	3
1.3	Light field acquisition technology	3
1.4	Motivation	5
1.5	Contributions	5
1.6	Outline	6
2	Background	8
2.1	Light field parameterization	9
2.2	Light field acquisition	9
2.3	Light field capabilities	11
2.3.1	Post-capture refocusing	13
2.3.2	Post-capture aperture adjustment	13

3	Extending aperture synthetically	16
3.1	Related work	17
3.2	Light field pre-processing	18
3.3	Light field registration	19
3.3.1	Rectification of sub-aperture images	20
3.3.2	Light field stitching	25
3.4	Experimental results	26
3.4.1	Synthetic aperture	29
3.4.2	Translation parallax	29
3.4.3	Disparity map range	29
3.5	Discussion and future work	31
4	Hybrid-sensing for resolution enhancement	33
4.1	Related work	33
4.2	Optical design	34
4.3	Spatially enhanced light field creation	36
4.3.1	Light field capture and pre-processing	37
4.3.2	Geometric registration	38
4.3.3	High-quality perspective formation	38
4.4	Experimental results	39

4.5	Discussion and future work	44
5	Controlled micro-shifts of light field sensor for resolution enhancement	45
5.1	Related work	46
5.2	Enhancing spatial resolution through micro-shifted light field sensor	47
5.2.1	The microscanning idea	47
5.2.2	Optical design and light field acquisition	48
5.3	Reconstructing high-resolution light field	51
5.3.1	Calibrating the lenslet grid	51
5.3.2	Light field registration	51
5.3.3	Non-uniform interpolation	52
5.3.4	Blind deblurring	52
5.4	Experimental results	53
5.4.1	Qualitative analysis	53
5.4.2	Quantitative comparison	56
5.4.3	Computation time	56
5.4.4	Effect of inaccurate shifts	58
5.5	Discussion and future work	59
6	Conclusions	60

List of Figures

1.1	A 17th century camera obscura illustration. Retrieved from cameraobscura.nz	1
1.2	Illustration demonstrating the basic difference between a traditional camera and a micro-lens array based light field camera.	4
2.1	Light field parameterization with two parallel planes. In each representation u and v serves as primary arguments. The last two arguments are parameterized by; (a) global coordinates of s and t , (b) angular coordinates θ and Θ representing the angle of ray after intersecting with uv plane, (c) local coordinates of s and t , are sometime also referred to as slope of the ray intersecting uv plane.	9
2.2	(a) Stanford camera array to create light field using an 8×16 cameras [1]. (b) A point in the scene is projected to three camera sensor of a camera array having independent optical elements.	10
2.3	(a) First generation Lytro camera [2]. (b) Optical diagram of a micro-lens array based light field camera.	11
2.4	Illustration of a light field being parametrized by actual u and s parallel planes and virtual u' and s' parallel planes.	12

2.5 Illustration of post-capture refocusing. (a, d, g) Optical diagram showing converging light rays at virtual image plane; (b, e, h) Pixels picked from raw lenslet marked with red to get refocused image after averaging marked points; (c, f, i) Three refocused images with each image having one depth in focus. 14

2.6 Illustration of post-capture aperture size adjustment. (a, d) Optical diagram demonstrating effect of placing a virtual aperture stop; (b, e) Region marked with red square shows the pixel region averaged to get the projected point; (c, f) Reconstructed image. . . 15

3.1 A lenslet image downloaded from a first generation Lytro camera. 19

3.2 (a) Middle sub-aperture image with two EPI lines marked. (b) EPI for the green line. Largest slope within the EPI is marked with a red line. (c) EPI for the blue line. Largest slope within the EPI is marked with a pink line. The largest slope among all EPIs is selected and used to compensate for the image center shifts. . . 20

3.3 Light field rectification and stitching illustrated with virtual cameras capturing sub-aperture images. The first light field is taken as the reference light field; and the second light field is rectified and stitched. The second light field images are rotated to compensate for the orientation difference of the light field cameras, scaled to compensate for the z-axis translations, and finally stitched to the first light field. 21

3.4 Extracted and depth clustered features. 24

3.5 Interpolation of sub-aperture images on a regular grid from rectified sub-aperture images. 24

3.6 Final light field obtained by merging of nine, six and ten light fields for first, second and third datasets respectively. (Right) Estimated sample locations and the resulting Delaunay triangulation for each dataset. Here each sub-aperture image denotes the image captured from different viewpoint in the scene. Whereas, its location in the multi-perspective view shows relative capturing position compared to other perspectives. 27

3.7 Epipolar plane image extension. (Top) Horizontal or vertical EPI lines marked for all three datasets. (Others) EPI for the single light field with their extended light field counterparts. 28

3.8 Comparison of refocusing and out-of-focus blurs at three different depths for single and extended light fields. 30

3.9 Translation parallax with the single light field and the extended light field. For first and second datasets: (Top) Leftmost sub-aperture image in the single light field and the extended light field; (Middle) Rightmost sub-aperture image in the single light field; (Bottom) Rightmost sub-aperture image in the extended light field. Whereas, for third dataset: (Top) Bottommost sub-aperture image in the single light field and the extended light field; (Middle) Topmost sub-aperture image in the single light field; (Bottom) Topmost sub-aperture image in the extended light field. 31

3.10 Disparity map comparison of single and extended light fields for Dataset 1. 32

4.1 (a) Proposed optical design; (b) Top view of the hardware setup based on the optical design. 35

4.2 Close-up of micro-lens images formed with different main lens aperture sizes (a) f/2.8, (b) f/4 and (c) f/8. The micro-lenses are f/4. 36

4.3 Illustration of the proposed light field resolution enhancement algorithm. Raw lenslet image is decoded using MATLAB toolbox [3] to create 4D plenoptic function. 37

4.4 Photometric registration of light field images and regular sensor image. (a) Middle sub-aperture light field image before photometric registration; (b) Middle sub-aperture light field image after photometric registration; (c) Homographically corrected image from regular sensor taken as the reference for photometric registration. 39

4.5 Spatial resolution enhancement of light field. (a) Low-resolution bilinearly interpolated light field captured by the light field sensor; (b) Spatially enhanced light field as a result of the proposed algorithm. 40

4.6 Spatial resolution comparison between original and spatially enhanced light fields for digital refocusing for Dataset 1. Using the shift-and-sum technique, light fields are focused at three different depths for both low-resolution and high-resolution light fields. . . 41

4.7 Spatial resolution comparison between original and spatially enhanced light fields for digital refocusing for Dataset 2. Using the shift-and-sum technique, light fields are focused at three different depths for both low-resolution and high-resolution light fields. . . 42

4.8 The quality of both light fields are compared after refocusing each of them to the current depth. At three different regions it is clearly visible that the lines are distinguishable pointing to high spatial resolution of reconstructed light field. 42

4.9 Quality of depth reconstruction is presented between (Left) low-resolution light field (Right) high-resolution light field; for two datasets. 43

5.1	Spatial resolution improvement of light field due to increased number of micro-lenses by merging two light field captures with a shift equal to the half of d_M (the distance between the centers of adjacent micro-lenses).	48
5.2	(Top-left) Proposed optical design; (Bottom-left) Optical hardware (zoomed-in side view); (Right) Optical hardware showing the translation stages.	49
5.3	Calibrated lenslet image with squares representing center pixel of each micro-lens (dotted circle). (Not drawn to scale.)	50
5.4	Combination of translated images to form high-resolution perspectives with the number of light field equal to: (a) 1; (b) 2; (c) 4; (d) 8; (e) 16.	50
5.5	Lenslet images with their magnified patches illustrating micro-lens distribution. (a) Low spatial resolution light field; (b) High spatial resolution light field by merging 16 light field captures.	54
5.6	Visual comparison of quality of middle perspectives between light fields generated with implementations of Dansereau <i>et al.</i> [3] and Cho <i>et al.</i> [4].	54
5.7	Visual comparison of quality of middle perspectives between light fields generated with different number of captures.	55
5.8	Comparison of post-capture refocusing demonstrated for light fields generated through (1) merging 16 captures, at three different depths, (2) Cho <i>et al.</i> [4] implementation, (3) Dansereau <i>et al.</i> [3] implementation, and (4) bicubic resizing.	56

5.9 Comparison of quality between high-resolution perspectives created with different number of captures based on probabilistic signal-to-noise ratio is presented by considering high-resolution perspective generated with 16 captures as a reference. 57

5.10 The time required to generate single high-resolution perspective for interpolation using different number of captures. 57

5.11 Comparison of different techniques to eliminate zig-zag artifacts due to inaccurate shifts. High-resolution perspectives are reconstructed using 16 light field captures, where: (Ideal) Actual shifts are applied with the assumption of zero translation error; (Cross Correlation) Shifts estimated using [5]; (SSD) Shifts estimated by minimizing sum-of-squared difference (SSD) between the images; (Averaging) By separately forming and averaging perspectives from even and odd rows of the micro-lens images, where the shifts are estimated using the SSD technique. 58

List of Tables

5.1	Translation amounts required to build a 4 times enhanced spatial grid in each direction, analogous to Figure 5.4e, with 16 light field captures.	50
-----	--	----

Chapter 1

Introduction

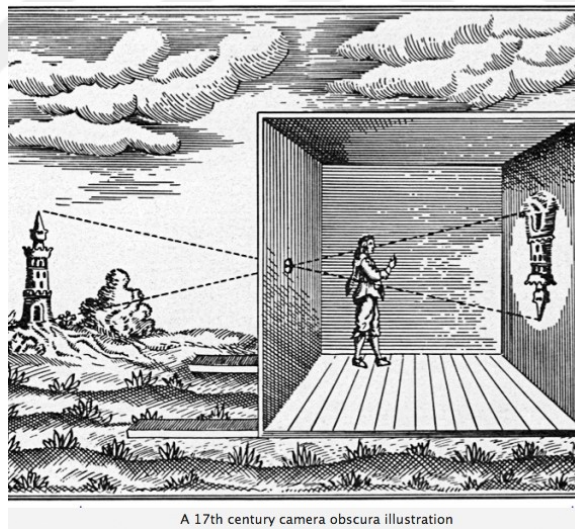


Figure 1.1: A 17th century camera obscura illustration. Retrieved from cameraobscura.nz

The word “camera” which is so common in our world today possesses a Latin origin, with its literal meaning of chamber or room. It originated from an experiment which was conducted in a dark room (*camera obscura*), with a pinhole at one wall of the room. The light rays reflected by the real-world scene pass through the pinhole to form a feeble inverted image over the other wall of the room. History records show that the first description for this principle was provided by the Chinese philosopher Mozi. Ibn al-Haytham performed many experiments based

on this principle and wrote down his observation in detail in the 11th century. Since that period, the technology of photography experienced a gradual change.

1.1 Traditional cameras

A traditional camera is principally composed of one or more optical elements and an imager which records the intensities spatially distributed over its surface. A lens is responsible for converging light rays reflected by the scene. If the imager is placed at a distance where light rays are converged by the lens, then it forms a sharp image over the imager.

The technology used to record images changed over the years. Till the middle of 20th century, images were recorded using light-sensitive silver halide crystals. Since the invention of charge coupled devices (CCDs), the technology is replaced by light sensitive semi-conductor elements for storing image in digital format.

A lens is a refractive surface usually made out of glass to change the direction of light. The amount of refraction depends on the refractive index and the angle of incident light to the lens surface. The surface of the glass is designed such that the light rays originating from a point source gets converged again at a point after passing through it, and this specific type of glass is known as a lens. In the past, it was difficult to make curved and smooth surface which limited the quality of the image. With recent technology, a lens is shaped almost perfectly to create better images.

1.2 Light field cameras

Contrary to a traditional camera which captures the intensity incident on the surface of the imager, a light field camera is an imaging device that captures the amount of light coming from different directions separately. Lippmann [6] first recorded the intensity of light rays from different directions by using a micro-lens array. The light field distribution in space and time was formulated by Gershun [7]. Adelson and Bergen [8] simplified Gershun's formulation and described light field with a five-dimensional representation (3 for position in space and 2 for direction). Levoy [9] and Gortler *et al.* [10] presented a four-dimensional representation of light field based on the assumption that there is no loss of energy in the direction of light ray.

In contrast to traditional cameras, light field cameras have angular resolution to describe the directional light amounts. Light field imaging has brought new applications and capabilities such as depth map estimation, refocusing, aperture control, perspective shifting and aperture coding.

1.3 Light field acquisition technology

There are two popular ways of capturing light field. One approach uses an array of cameras [9], [11] to capture a set of two-dimensional images. Here, the number of cameras determines angular resolution; and, the spatial resolution of the light field perspective image is determined by the spatial resolution of a camera in the array. An example of multi-camera array based system was presented by a team of researchers in Stanford University [1]. The array is used for variety of application including high dynamic range imaging, seeing through partially occluded environments, and resolution enhancement.

The other popular approach is to use a micro-lens array (MLA) in front of an image sensor [12], [13]. This method gained popularity due to low cost and compact area it takes. The initial idea to capture light field using MLA was

presented by Gershun [7]. Ng [12] built a compact light field camera design by placing the sensor at the focal length of micro-lens array (micro-lens focal length is in micrometers). The design is eventually commercialized as the Lytro camera [2]. There are several other micro-lens array based designs which use micro-lens array at some intermediate image plane such as Raytrix [14] and Plenoptic camera 2.0 [13]. As shown in Figure 1.2, in comparison with a traditional camera, a light field sensor does not accumulate the light intensities at the image plane, but it provides a passage to light rays at that point so that light can further traverse and the intensities of these light rays are recorded separately by a sensor placed at some distance from the focus plane.

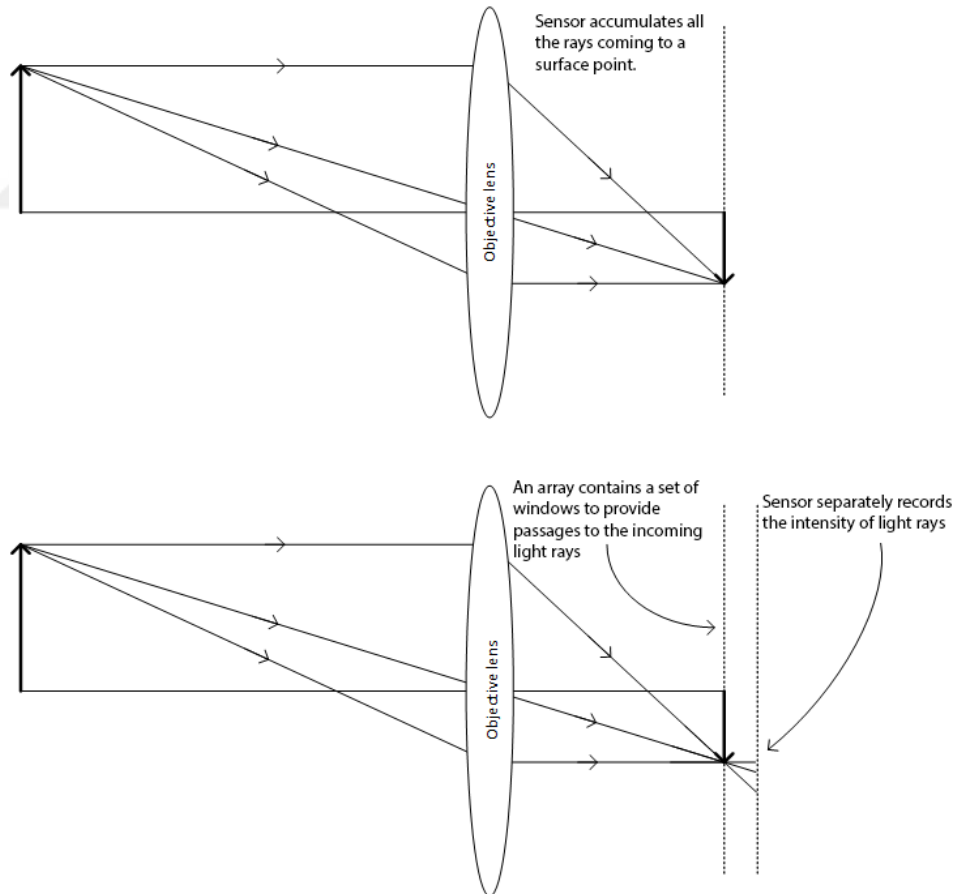


Figure 1.2: Illustration demonstrating the basic difference between a traditional camera and a micro-lens array based light field camera.

There are also other light field capture systems, such as coded mask [15], lens array [16], camera moved on a gantry [17], and kaleidoscope-like optics [18].

1.4 Motivation

Micro-lens array based light field cameras pose some serious advantage over multi-camera based light field systems, such as cost effectiveness and compactness; however, these advantages are achieved at the cost of certain limitations affecting the image quality and performance.

One limitation of MLA based light field cameras is small aperture size. With a large aperture, light field can be utilized for better depth estimation or seeing through the occluded regions. Moreover, with a larger aperture, higher angular range can be achieved.

The second major issue with micro-lens array based light field cameras is low spatial resolution. This is due to the trade-off between angular resolution and spatial resolution; and many light field cameras sacrifice spatial resolution to gain angular information. For instance, in the first generation Lytro camera [2], the shared resolution of sensor limits the spatial resolution to less than 0.15 megapixels for achieving an angular resolution of 11x11 perspectives¹. Such a spatial resolution is low for many applications; hence it limits the use of light field cameras.

1.5 Contributions

The main contributions in this thesis are as follows:

Light field stitching for extending synthetic aperture - In this work, a technique is presented to overcome the limited aperture size by means of multiple light field captures in random orientations. Multiple light field captures are first transformed and then stitched to synthetically extend the aperture. This idea is

¹The open-source decoding software presented by Dansereau [3] produces 380x380 pixel resolution images. Lytro's proprietary software, which incorporates some sort of interpolation/enhancement algorithm, processes the raw lenslet image to achieve a spatial resolution of 1.2 mega-pixels.

published as a conference paper [19].

Hybrid-sensor high-resolution light field imaging In this work, features of a traditional camera sensor and a light field sensor are blended together to achieve a high spatial resolution light field image. The optical system for this idea is designed with an excess of a beam-splitter (which splits the projected rays into two), which makes this design cost-effective. After a motion estimation step the high-resolution image obtained from the traditional camera is registered to the low-spatial-resolution perspective images of the light field. At the end, high-spatial resolution light field perspective images are obtained. The resulting light field contains high frequency features, which improves epipolar plane image (EPI) quality; hence resulting in better performance with EPI based depth estimation algorithms. The idea is published as a conference paper [20].

Light field super resolution through controlled micro-shifts of light field sensor - In this work, resolution enhancement is achieved through controlled shift of light field sensor and multiple light field captures. The resolution enhancement factor is equal to the number of light fields acquired. The idea is based on the super-resolution enhancement of low spatial resolution images obtained by traditional cameras. In the context of light field imaging, the technique has been used for the first time. Implementation of this idea has proved that it has the potential to solve the problem of limited spatial resolution of light field cameras by incorporating a linear motion creating actuators. This idea is currently in the review process for a journal publication.

1.6 Outline

The light field technology is relatively new, hence the document expects basic understanding of this topic from the reader before proceeding to the contributions of the thesis. Therefore, **Chapter 2** gives some brief insight on how the technology works. It initially forms the basis by giving some brief history about the concept of light field. It then describes that how a light field can be effectively

represented parametrically. The technologies which are currently often used to capture light field is also described in that chapter. Finally, it explains some post-capture capabilities of light field.

In **Chapter 3**, a method to overcome the narrow aperture problem of a light field camera is provided. The methodology along with the effect of increasing the aperture is discussed in that chapter. In the end, results are presented demonstrating the effect of increased aperture size.

In **Chapter 4** and **Chapter 5**, two solutions are provided which can computationally deal with the low-spatial resolution problem of light field with some hardware modification. In Chapter 4, the outcome of two different type of imaging sensors are intertwined to achieve a superior light field image in terms of spatial resolution. In Chapter 5, a different approach is provided to tackle the issue of low spatial resolution by shifting the light field sensor on the image plane in micrometer levels. These images are combined at a later stage to have a higher spatial resolution light field. The achieved spatial resolution is then compared against different methods.

Chapter 2

Background

A two dimensional image recorded by a traditional camera, is merely a projection of light on the sensing surface reflected by the 3D world. Whereas, light is a higher-dimensional quantity. An eye gathers the light spread from the 3D world. The rays reflected from the surface in the scene converges after passing through lens to form an image over retina. In 1991, Gershun [7] proposed a seven-dimensional representation of light field which consists of one dimension for time, three dimensions for space, two dimensions for direction and one dimension for frequency. They termed this representation as the plenoptic function. Levoy [9] and Gortler *et al.* [10] presented a concise four-dimensional representation of this function by constraining that there is no loss of energy in the direction of light ray. Their representation eased up image rendering process as the reduced dimensionality reduces the memory requirement. In this representation, time dimension is discarded as the scene is static; and the frequency dimension which is used to represent wavelength is replaced by a three-color (RGB) notation in a digital system. Moreover, by constraining over the assumption of no loss of energy in the direction of light ray propagation, the five-dimensional representation (3 for space and 2 for direction) reduces to four-dimensional representation (2 for space and 2 for direction) for each color channel.

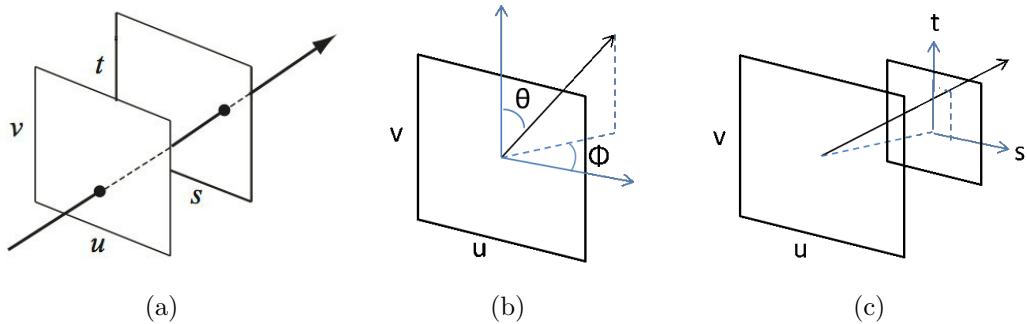


Figure 2.1: Light field parameterization with two parallel planes. In each representation u and v serves as primary arguments. The last two arguments are parameterized by; (a) global coordinates of s and t , (b) angular coordinates θ and Φ representing the angle of ray after intersecting with uv plane, (c) local coordinates of s and t , are sometime also referred to as slope of the ray intersecting uv plane.

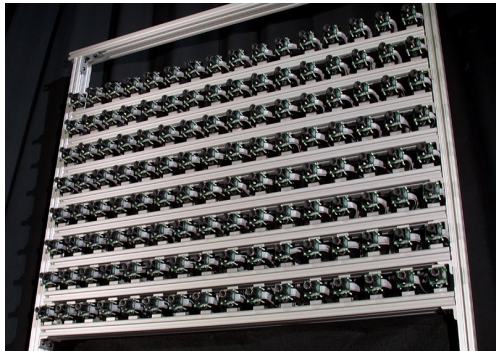
2.1 Light field parameterization

Parametrization of light field is important to index each light ray approaching the optical system. The reduction of dimensionality to a four-dimensional function for light field representation results in several possible representation to parameterize a four-dimension light field. The intersection of a ray in space with two parallel planes can completely parameterize ray in terms of point of intersection on both planes. Some models are shown in Figure 2.1, describing ways to parameterize light field.

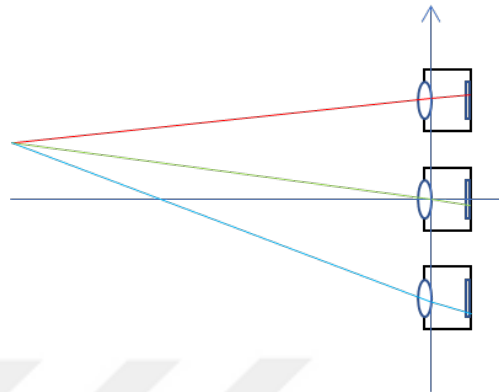
2.2 Light field acquisition

There are two popular ways of capturing light field. One is to use an array of cameras [9], [11] and the other is to use a micro-lens array (MLA) in front of an image sensor [12], [13]. There are also other light field capture systems, such as coded mask [15], lens array [16], camera moved on a gantry [17], and kaleidoscope-like optics [18].

To capture light field, the simplest approach is to use an array of cameras. Such



(a)



(b)

Figure 2.2: (a) Stanford camera array to create light field using an 8×16 cameras [1]. (b) A point in the scene is projected to three camera sensor of a camera array having independent optical elements.

a camera setup, developed by Stanford University [1], is shown in Figure 2.2. The number of cameras in the grid represents the angular resolution, such as, a grid of 8×16 grid of camera array will give an angular resolution of 8×16 . Typically, the baseline between adjacent cameras is wider compared other light field captures, which results in a discrete blur when all the perspectives are combined together. Although the quality of the light field produced through this setup is superior to every other technique and suitable for dynamic scenes, it typically requires large area and is not a cost-effective solution to capture light field.

In contrast to the camera array approach, a micro-lens array based light field acquisition technique is an effective solution due to its cost efficiency and compactness. Ng [12] presented a compact light field camera design by placing a micro-lens array at the focal plane of the main lens, and an image sensor separated from the micro-lens array by an amount equal to the focal length of the lenslets in the micro-lens array. The design is later commercialized as the “Lytro” camera, which is shown in Figure 2.3. The first generation Lytro utilizes a hexagonally gridded micro-lens array which gives a spatial resolution of 380×380 pixels approximately. Moreover, the baseline between adjacent virtual cameras in the resulting array is very small, which limits depth estimation capability of the light field.

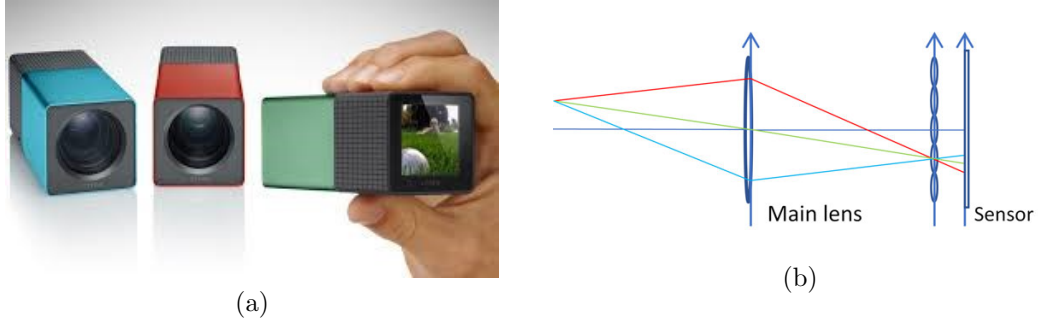


Figure 2.3: (a) First generation Lytro camera [2]. (b) Optical diagram of a micro-lens array based light field camera.

2.3 Light field capabilities

With a traditional camera, the optical parameters, such as focus and aperture size, are preset before an image is captured. Light field cameras offer post-capture capability to refocus, adjust the aperture size, and change the perspective by capturing directional information of light rays. To understand the concept of digital post-capture capabilities of light field camera we need to derive a mathematical relationship between actual light field and synthetic light field that have synthetic parallel planes placed at different position when compared with their actual counterparts.

In Figure 2.4, the synthetic light field L' is parameterized by the synthetic u' , v' and s' , t' planes. In the diagram, F is taken as the separation between actual parallel planes, α is the ratio of the positions of virtual s' and actual s planes with respect to actual u plane, while β is the ratio of the positions of virtual u' and actual u planes with respect to actual s plane. In this figure, it can be seen that a ray is intersecting these parallel planes at u' and s' . (Only 2 dimensions out of 4 dimensions are shown to simplify the understanding.) The irradiance of the intersecting ray then can be written in the form [21]:

$$\bar{E}(s', t') = 1/D^2 \iint L'(u', v', s', t') A(u', v') \cos^4 \theta du' dv', \quad (2.1)$$

where D is the separation between the two synthetic planes, A is an aperture multiplier function that becomes one within the aperture opening and zero outside it, and θ is the angle of incidence that ray (u, v, s, t) makes with the film plane.

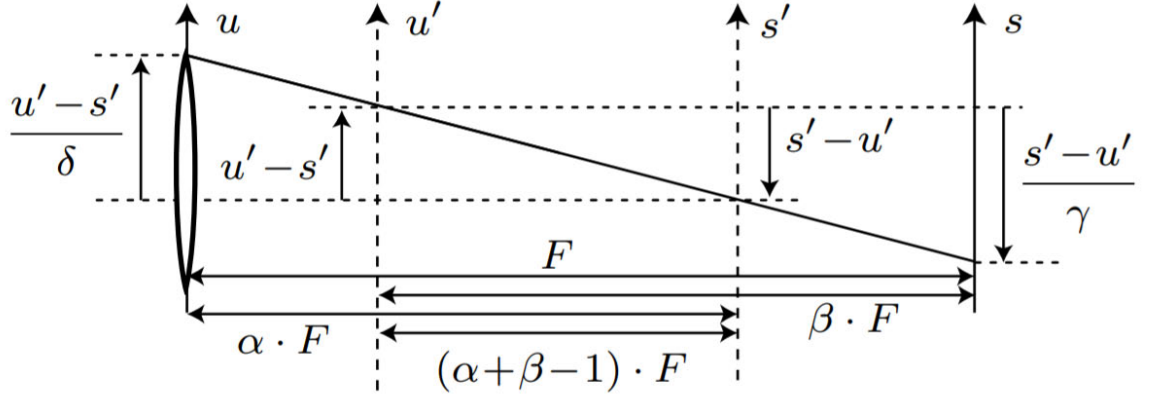


Figure 2.4: Illustration of a light field being parametrized by actual u and s parallel planes and virtual u' and s' parallel planes.

To simplify this equation, a paraxial approximation is applied to the above equation. Moreover, D^2 term is neglected by setting it equal to 1 to obtain

$$\bar{E}(s', t') = \iint L(u', v', s', t') A(u', v') du' dv'. \quad (2.2)$$

The diagram in Figure 2.4 forms the understanding to express equation in terms of the captured light field $L(u, v, s, t)$. The illustration will help in expressing a mathematical relationship between L' and L . In addition, we can define $\gamma = (\alpha + \beta - 1)/\alpha$ and $\delta = (\alpha + \beta - 1)/\beta$ for notational convenience. The ray intersecting u' and s' also intersects the u plane at $s' + (u' - s')/\delta$ and the s plane at $u' + (s' - u')/\gamma$. Thus,

$$L'(u', v', s', t') = L\left(s' + \frac{(u' - s')}{\delta}, t' + \frac{(v' - t')}{\delta}, u' + \frac{(s' - u')}{\gamma}, v' + \frac{(t' - v')}{\gamma}\right). \quad (2.3)$$

Combining these equations, one can obtain the synthetic photography equation, which is used as an image formation model [21]:

$$\bar{E}(s', t') = \iint L\left(s' + \frac{(u' - s')}{\delta}, t' + \frac{(v' - t')}{\delta}, u' + \frac{(s' - u')}{\gamma}, v' + \frac{(t' - v')}{\gamma}\right) A(u', v') du' dv'. \quad (2.4)$$

The above equation can be used to model the effect of post-capture capabilities

of light field after molding the form of this equation. Therefore, in the upcoming sections, two of the important light field capabilities are modeled and then proved using the above equation.

2.3.1 Post-capture refocusing

The rays approaching to the light field camera are parametrized as $L(u, v, s, t)$, where u and v are the angular coordinates, while s and t are spatial coordinates. Keeping u and v fixed (picking up pixels from same u and v locations of each micro-lens), will form a perspective. Using image formation model we can generate image at virtual movable st plane. In refocusing, only the synthetic film plane moves (i.e. $\beta = 1$), and we use a full aperture (i.e. $A(u', v') = 1$). In this case $\delta = \alpha$ and $\gamma = 1$, and the synthetic photography equation simplifies to

$$\bar{E}(s', t') = \iint L\left(u', v', u' + \frac{(s' - u')}{\gamma}, v' + \frac{(t' - v')}{\gamma}\right) du' dv'. \quad (2.5)$$

Examining this equation reveals the important observation that refocusing is conceptually just a summation of shifted versions of the images that form through pinholes (fix u' and v' and let s' and t' vary) over the entire uv aperture. In quantized form, this corresponds to shifting and adding the sub-aperture images, which is the technique used (but not physically derived) in previous papers [22, 23]. Figure 2.5 shows the refocused images at three different depths.

2.3.2 Post-capture aperture adjustment

An aperture is an opening of a lens through which light can enter into an optical system. For a fixed focal length, an aperture size can be used to describe the cone of angle of the bundle of rays that focuses on image plane. A small sized aperture will allow only the bundle of rays with limited cone angle to pass through the lens, resulting in an overall sharp image. Whereas, a large aperture size, will produce a sharp image for the object in focus and the parts from the other depth gets

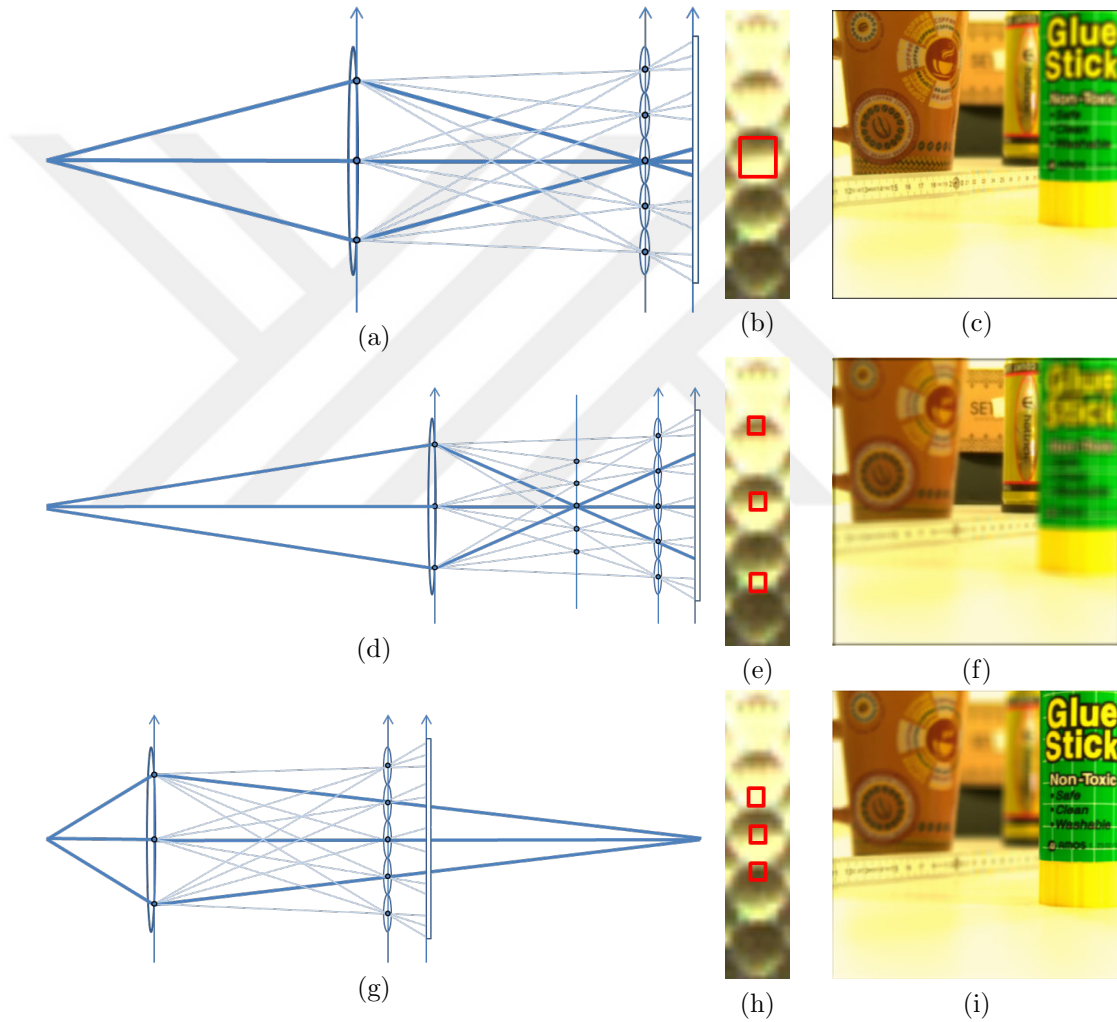


Figure 2.5: Illustration of post-capture refocusing. (a, d, g) Optical diagram showing converging light rays at virtual image plane; (b, e, h) Pixels picked from raw lenslet marked with red to get refocused image after averaging marked points; (c, f, i) Three refocused images with each image having one depth in focus.

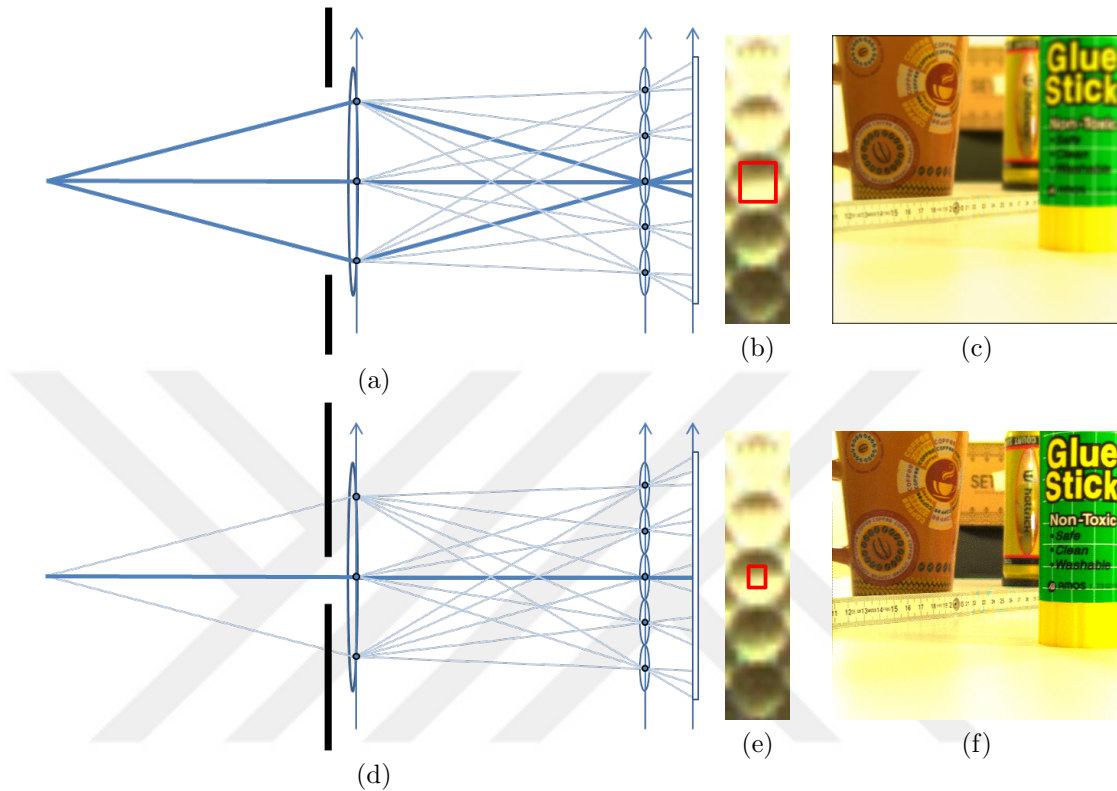


Figure 2.6: Illustration of post-capture aperture size adjustment. (a, d) Optical diagram demonstrating effect of placing a virtual aperture stop; (b, e) Region marked with red square shows the pixel region averaged to get the projected point; (c, f) Reconstructed image.

blurred proportional to their distances from the object in focus. There are variety of applications which can be achieved by changing the aperture size. Such as, decreasing the aperture size can help reduce the effect of optical aberration. This concept is evident by the image formation model, that for the reduced aperture size $A(u', v')$ becomes zero for the extreme coordinates of u and v , which will effect the contribution of the rays in the overall integration approaching from extreme coordinates of u' and v' . Figure 2.6 demonstrates that by reducing the aperture size, depth of field is increased. By reducing the aperture size the cone angle of the bundle of ray becomes narrower, resulting in an overall sharp image.

Chapter 3

Extending aperture synthetically

By combining multiple light fields, it is possible to obtain new capabilities and enhancements, and even exceed physical limitations, such as spatial resolution and aperture size of the imaging device. In the previous chapters, light field technology is explained and some of its major problems are pointed which limits the performance of this technology. One of them is the limited aperture size. In this chapter, the idea of extending aperture synthetically by capturing multiple light fields is introduced. An algorithm to register and stitch multiple light fields is presented. The regularity of the spatial and angular sampling in light field data is utilized, and some of the techniques that are developed for stereo vision systems have been extended to light field data in this research. Such an extension is not straightforward for a micro-lens array (MLA) based light field camera due to extremely small baseline and low spatial resolution. By merging multiple light fields captured by an MLA based camera, larger synthetic aperture is obtained, which resulted in improvement in light field capabilities, such as increased depth estimation range/accuracy and wider perspective shift range.

3.1 Related work

Registration of multiple light field captures has recently been addressed in a few publications. In [24], a method for creating panoramic light fields is presented. The method is based on projecting two-plane parameterized light fields on a cylindrical coordinate system. The method is limited to rotational motion between light fields; thus, the light field camera must be rotated around its focal point. This requires fixing the camera on a tripod and precise alignment of the rotational center of the tripod with the focal point.

The method presented in [25] is not restricted to rotation around the optical center, and can handle translation as well rotation. It is based on transforming the light field ray parameters to Plücker coordinates, which results in a projective transformation, named ray-space motion matrix (RSMM), between two light fields. The SIFT features are extracted from sub-aperture views to determine the ray correspondences; and the RSMM is estimated from the ray correspondences. It is reported that the method requires large overlap between the light fields to have enough ray correspondences and even with large overlaps rays may not match exactly due to undersampling. This may cause imperfect RSMM estimates, and a graph-cut based refinement step is utilized. One drawback of the method is the high computational cost: The average time to stitch a pair of light fields (captured by a Lytro camera) is about 20 minutes (on a PC with Intel i7 CPU with 64GB memory). Another Plücker coordinate system based approach is presented in [26]. Ray correspondences are also determined using SIFT features; and the optimization is done based on [27].

It should be noted that creating a panoramic light field requires the camera to be rotated around the optical center as in [24]. When the translation of the camera is allowed, an attempt to create panoramic light field may suffer from “ghosting artifacts” due to translation parallax [25]. Because of this fundamental issue, it may be a better idea to generate extended light field aperture instead of attempting to create panoramic view when there is translation of light field camera.

In this research, multiple light fields are registered and merged to obtain a light field with larger synthetic aperture. Different from the previous methods, this registration approach is based on the epipolar geometry of light field data. While epipolar geometry based registration has been studied extensively for structure from motion, the application for light field data is not straightforward when the data is captured with a micro-lens array based camera, such as the Lytro, which has low spatial resolution, low signal-to-noise ratio, and narrow baseline between the sub-aperture images. This approach successfully works with such data.

3.2 Light field pre-processing

For the experiments, a first generation Lytro camera [2] is used. Figure 3.1 shows a raw lenslet image downloaded from the camera. There are several tools developed to decode the raw lenslet image [4], [28] and [3]. A third party toolbox [3] developed in MATLAB is used to decode light field from raw lenslet image. A 9×9 array of sub-aperture images are picked for further processing, although the actual number of perspectives were 11×11 . The extreme perspectives are discarded as most of them contain low SNR and highly vignetted images.

Before proceeding towards registration of light field stage there are two essential pre-processing steps to be performed on decoded images. The first one is vignetting correction. The intensity of sub-aperture images decreases from middle to side perspectives due to vignetting. A histogram-based photometric mapping is used to match the colors of each perspective with the middle perspective image [29]. For a robust mapping, filtering of each perspective with a Gaussian filter (of size 5×5 and with standard deviation 0.6) is employed to improve SNR.

The second pre-processing step is image center correction. The optics of the Lytro camera focuses at some finite depth. The architecture of the light field camera makes the disparity among different perspectives zero for that depth in the scene. The overall effect appears like disparities are offset to some specific value. This is clearly seen in the epipolar plane images (EPIs) in Figure 3.2,



Figure 3.1: A lenslet image downloaded from a first generation Lytro camera.

where the EPIs include lines with slope larger than 90 degrees (measured from the positive x-axis in the counter-clockwise direction). The largest slope would be 90 degrees if the array were focused at the farthest depth in the scene. Furthermore, it is not guaranteed that the array focuses at the same depth from one light field capture to another. To have the same common reference plane among all light fields, which will be used during the stitching process, all sub-aperture images in a light field are translated to ensure focusing at the farthest depth in the scene. The EPI slope based approach [30] is used to estimate the translation amount: The slope of all the line in the epipolar images are estimated using the Hough transform [31]; the largest slope is determined among all EPIs, and each sub-aperture image is translated accordingly. (The process is repeated for horizontal and vertical directions.)

3.3 Light field registration

The light field registration approach consists of rectification and stitching steps. During rectification, all sub-aperture images are compensated for rotation and translation so that they are on the same plane. During stitching, the rectified

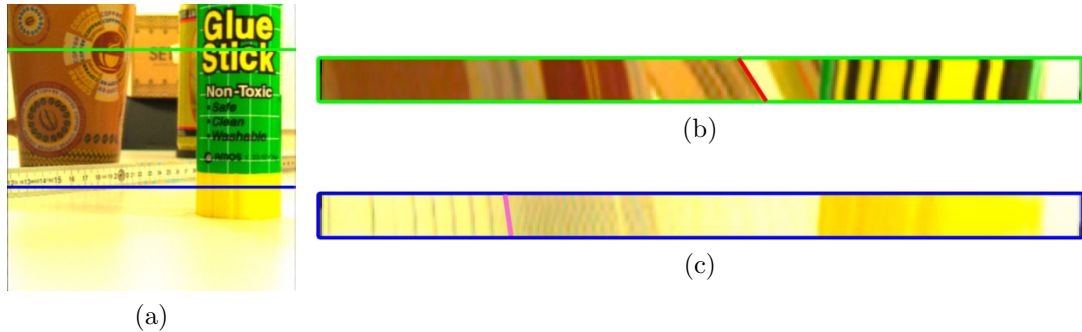


Figure 3.2: (a) Middle sub-aperture image with two EPI lines marked. (b) EPI for the green line. Largest slope within the EPI is marked with a red line. (c) EPI for the blue line. Largest slope within the EPI is marked with a pink line. The largest slope among all EPIs is selected and used to compensate for the image center shifts.

sub-aperture images are merged into a single light field. These steps are detailed in the following sections.

3.3.1 Rectification of sub-aperture images

A light field camera can be modeled as an array of virtual cameras, each capturing a sub-aperture (i.e., perspective) image. In case of the Lytro camera, the regularity of the micro-lens array in front of the sensor results in sub-aperture images captured by virtual cameras with regular spacings and identical orientations. In Figure 3.3, an illustration with two virtual camera arrays as the captured light fields is provided. (The algorithm for stitching two light fields is explained; the process is repeated for each additional light field.) The sub-aperture images of the second light field are rotated and translated with respect to the first light field sub-aperture images. While the translations differ, the rotation amount between a virtual camera of the first light field and a virtual camera of the second light is identical. First, the orientations of the second light field sub-aperture images are corrected. After orientation correction, the scale is corrected to place both light fields onto the same plane.

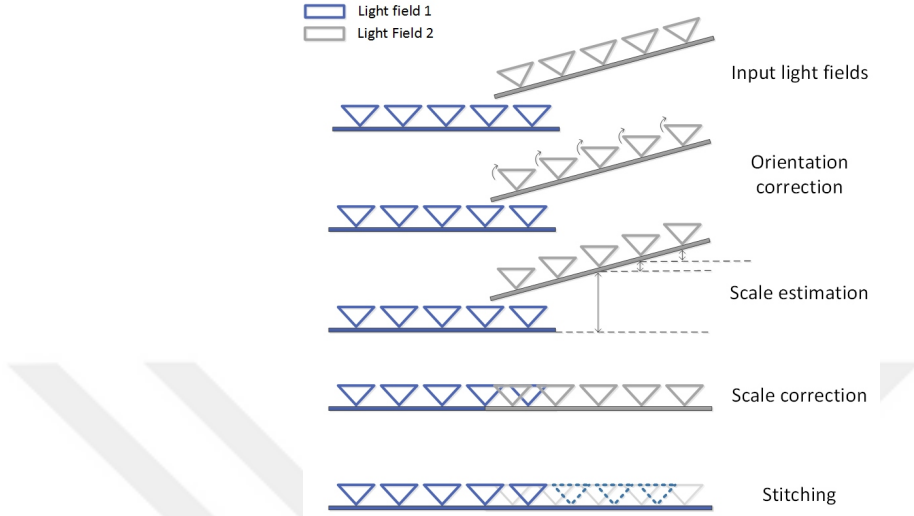


Figure 3.3: Light field rectification and stitching illustrated with virtual cameras capturing sub-aperture images. The first light field is taken as the reference light field; and the second light field is rectified and stitched. The second light field images are rotated to compensate for the orientation difference of the light field cameras, scaled to compensate for the z-axis translations, and finally stitched to the first light field.

3.3.1.1 Orientation correction

The orientation difference is estimated through the fundamental matrix of any sub-aperture image pair from the first and second light fields. The middle sub-aperture images of each light field are used to estimate the fundamental matrix through feature correspondences as done in traditional stereo imaging systems [32]. Harris corner features [33] are extracted in the middle sub-aperture image of the first light field and use the Kanade-Lucas-Tomasi (KLT) algorithm [34] to obtain the correspondences in the middle sub-aperture image of the second light field. The fundamental matrix is then estimated after moving the outliers from the correspondences.

To clarify further, suppose that the corresponding feature coordinates are (u_i, v_i) and (u'_i, v'_i) in the middle sub-aperture image of the first light field and the middle sub-aperture image of the second light field. Outliers from the correspondences are removed through RANSAC technique such that the fundamental matrix equation, $[u_i, v_i, 1]F[u'_i, v'_i, 1]' = 0$, where F is the fundamental matrix, is

satisfied. After the outliers are removed; the fundamental matrix is estimated that minimizes the re-projection error using the gold standard technique [32]. Using the intrinsic camera matrix K , which is formed using the camera parameters (i.e., pixel pitch and focal length) available in the light field metadata, the essential matrix $E = K^T F K$ is calculated. The essential matrix is then decomposed to obtain the rotation matrix [35]. Specifically, the essential matrix is first decomposed using singular value decomposition (SVD):

$$E = U \Sigma V^T, \quad (3.1)$$

where U and V are orthonormal matrices and $\Sigma = \text{diag} \{ \sigma_1, \sigma_2, \sigma_3 \}$ is a diagonal matrix, with σ_1, σ_2 and σ_3 being the diagonal elements. For an essential matrix, the first two diagonal elements must be identical and the third element must be equal to zero. To impose this condition, a revised essential matrix is constructed with an updated diagonal matrix $\Sigma = \text{diag} \{ (\sigma_1 + \sigma_2)/2, (\sigma_1 + \sigma_2)/2, 0 \}$, which is optimal in terms of the Frobenius norm [36]. The new essential matrix is decomposed again using SVD: $E = U \Sigma V^T$, and the rotation matrix R is calculated as:

$$R = U W V^T, \quad (3.2)$$

where W takes two possible versions [36]:

$$W = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.3)$$

Among the two viable solutions for the rotation matrix, only one is physically realizable, which is chosen such that the reconstructed points have positive depths [36].

The estimated rotation matrix is then applied to every sub-aperture image of the second light field to correct for the orientation using the homographic transformation $[\alpha u'', \alpha v'', \alpha]^T = K R K^1 [u', v', 1]^T$, where (u', v') are the pixel coordinates in a sub-aperture image and (u'', v'') are the transformed coordinates.

3.3.1.2 Scale estimation and correction

After the orientation correction, compensation for the z-axis translations (i.e., translations orthogonal to the first light field image plane) within the second light field and between the first and second light fields is required. The effect of these translations is scale change between the images. The scale of each sub-aperture image from the second light field needs to be calculated separately.

Within-light-field scale estimation: Because the scale is fixed between consecutive pairs of the second light field sub-aperture images, the scale is estimated between every consecutive pair within the second light field and take the geometric mean to have a robust estimate. The scale estimation is again based on feature correspondences. The same procedure (Harris corner detection followed by KLT based feature tracking) to obtain the feature correspondences is followed. To properly estimate the scale, the features from the same depth should be used. The histogram of distances between the correspondences reveal the number of depths available in the scene. The number of depth clusters in the scene are extracted according to the Silhouette's criterion [37] through fitting mixture of Gaussians over the distribution. Features are assigned to a cluster based on their Euclidean distances to the cluster centroids. (The extracted and clustered features from the light field given in Figure 3.1 are shown in Figure 3.4 as an example.) To estimate the scale, features from any depth cluster can be used; the features from the farthest depth cluster are used. Similarity transformation is fitted to the feature correspondences between a pair of sub-aperture images to get the scale between the pair.

Between-light-field scale estimation: The scale between the light fields are estimated by applying the same procedure described above on the middle sub-aperture images of the first and second light fields.

Scale correction: The estimated within-light-field scales and between-light-field scale are multiplied to obtain the overall scale of each sub-aperture image of the second light field. These scales are then applied to bring all sub-aperture images on the same plane.

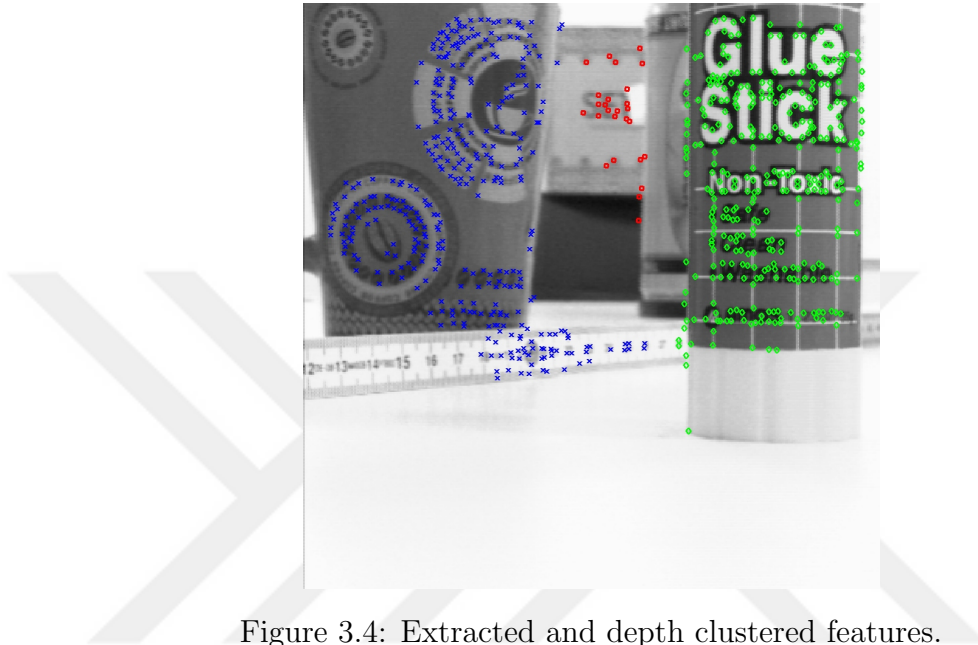


Figure 3.4: Extracted and depth clustered features.

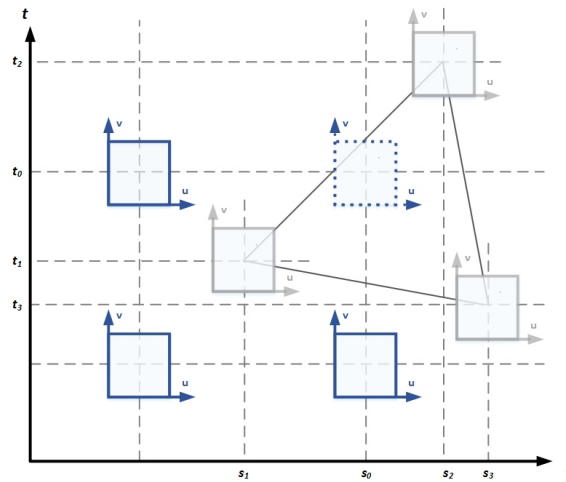


Figure 3.5: Interpolation of sub-aperture images on a regular grid from rectified sub-aperture images.

3.3.2 Light field stitching

The last step is to merge the light fields into one. While the sub-aperture images are now all rectified (rotated and scaled), the translation amounts are yet to be determined. The feature correspondence based approach is used again to determine translations. Using feature correspondences, the within-light-field translation amounts are first estimated between two consecutive sub-aperture pairs. Since the translation amount is fixed between two consecutive pairs, the translation between every pair is estimated and then averaged to have a robust estimate. The translation between the light fields is then estimated using the middle sub-aperture images. Combining within-light-field and between-light-field translations, the translations for every sub-aperture image are obtained. The translation amounts may not correspond to regular grid locations; to obtain a light field on a regular grid, therefore interpolation is needed. The Delaunay triangulation technique for interpolation is used. As shown in Figure 3.5, the irregular positions of the light fields are triangulated to obtain new sub-aperture images at uniform grid positions using pixel-by-pixel weighted sum of neighboring sub-aperture images: Referring to Figure 3.5, suppose that (s_0, t_0) is the grid position where the sub-aperture image has to be estimated, and (s_i, t_i) with $i = 1, 2, 3$ are the locations where the light field sub-aperture images $I(u, v, s_i, t_i)$ are recorded. If (s_0, t_0) is equal to one of the recorded sample location (s_i, t_i) , then the sub-aperture image is directly set to the recorded subaperture image at that location. Otherwise, the sub-aperture image $I'(u, v, s_0, t_0)$ is interpolated as a weighted sum of recorded images $I(u, v, s_i, t_i)$:

$$I'(u, v, s_0, t_0) = \sum_{i=1}^3 \lambda_i I(u, v, s_i, t_i), \quad (3.4)$$

where these weights λ_i are determined using barycentric scheme [38], through the following set of equations: $s_0 = \sum_{i=1}^3 \lambda_i s_i$, $t_0 = \sum_{i=1}^3 \lambda_i t_i$, and $\sum_{i=1}^3 \lambda_i = 1$.

3.4 Experimental results

In this section, experimental results for three datasets captured with a first generation Lytro camera are provided. The light field toolbox of [3] is used to decode the light fields. All implementations are done in MATLAB, running on an i5 PC with 12 GB RAM.

The first dataset consists of nine light fields in which camera movement is in horizontal direction. The second dataset consists of six light fields, where the camera is moved in horizontal direction along with a rotation motion around z-axis. The third dataset includes ten light fields, where the camera is arbitrarily moved in both horizontal and vertical directions. During all these captures camera is moved with hand, which means that the motion fields can contain shift and rotation in any axis.

The pre-processing time per light field is about 16 seconds, and the rectification time per light field is about 10 seconds. The stitching time depends on the final grid size. The extended light field for the first dataset has a final grid of size 9×24 . The extended light field for the second dataset has a final grid of size 26×33 . The extended light field for the third dataset has a final grid of size 13×28 . The stitching times are 140, 180 and 300 seconds for the first, second and third datasets, respectively. The extended light fields for the three datasets with their corresponding sub-aperture locations and Delaunay triangulations used in the interpolation are provided in Figure 3.6.

The extension of light field for all three datasets is also demonstrated by presenting the extension of EPI range in Figure 3.7. The EPI demonstrates the extension of the aperture; the straightness of the feature lines in the EPI indicates the correctness of the registration process.

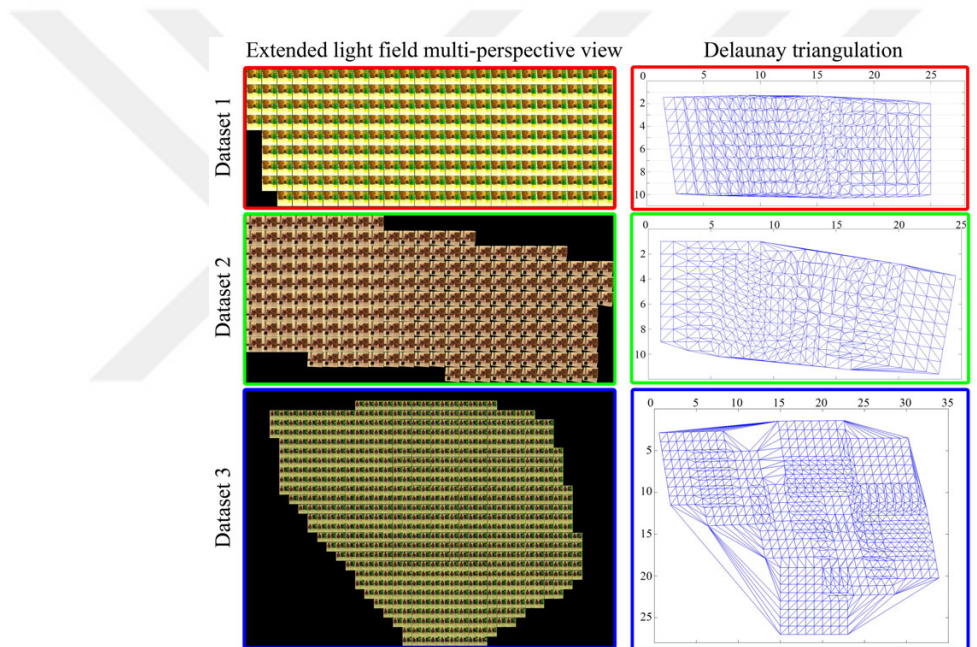


Figure 3.6: Final light field obtained by merging of nine, six and ten light fields for first, second and third datasets respectively. (Right) Estimated sample locations and the resulting Delaunay triangulation for each dataset. Here each sub-aperture image denotes the image captured from different viewpoint in the scene. Whereas, its location in the multi-perspective view shows relative capturing position compared to other perspectives.

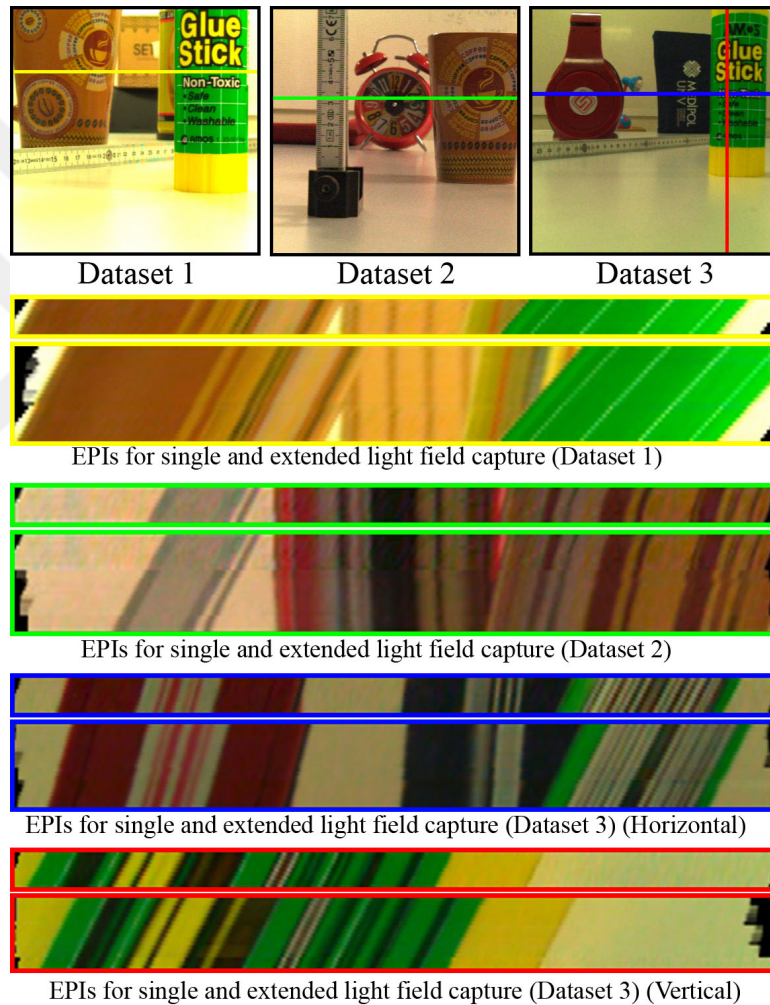


Figure 3.7: Epipolar plane image extension. (Top) Horizontal or vertical EPI lines marked for all three datasets. (Others) EPI for the single light field with their extended light field counterparts.

3.4.1 Synthetic aperture

One of the features of light field photography is the ability to digitally change focus after capture. With a larger aperture, the refocusing effect becomes more dramatic as the blur in the out-of-focus regions are larger. In Figure 3.8, the light fields are focused at different depths using the shift-and-sum technique [9]. The sharpness of the images in the focused regions indicates that the light fields are properly registered. The amount of blur in the out-of-focus regions is larger due to the extended aperture. It can also be noticed that the direction of the blur reflects the extension of the aperture. For example, in Figure 3.8, for the first and second datasets the blur is more in the horizontal direction, while for the third dataset, the blur is more in the vertical direction.

3.4.2 Translation parallax

With the extension of aperture, the baseline between the extreme sub-aperture images of the extended light field is also increased. The effect can be clearly seen by comparing the extreme sub-aperture images of a single light field and extended light field. In Figure 3.9, for first and second datasets, horizontal translation parallax for the single and extended light fields are shown: The top image is the leftmost sub-aperture image in the single light field and the extended light field, the middle image is the rightmost sub-aperture in the single light field, and the bottom image is the rightmost sub-aperture in the extended light field. The increase in translation parallax is visible when these images are compared. Similarly, in the same figure for third dataset, the vertical translation parallax for single and extended light fields is compared.

3.4.3 Disparity map range

MLA based light field cameras, such as Lytro, have narrow baseline between the sub-aperture images. This limits the depth map estimation range and accuracy.

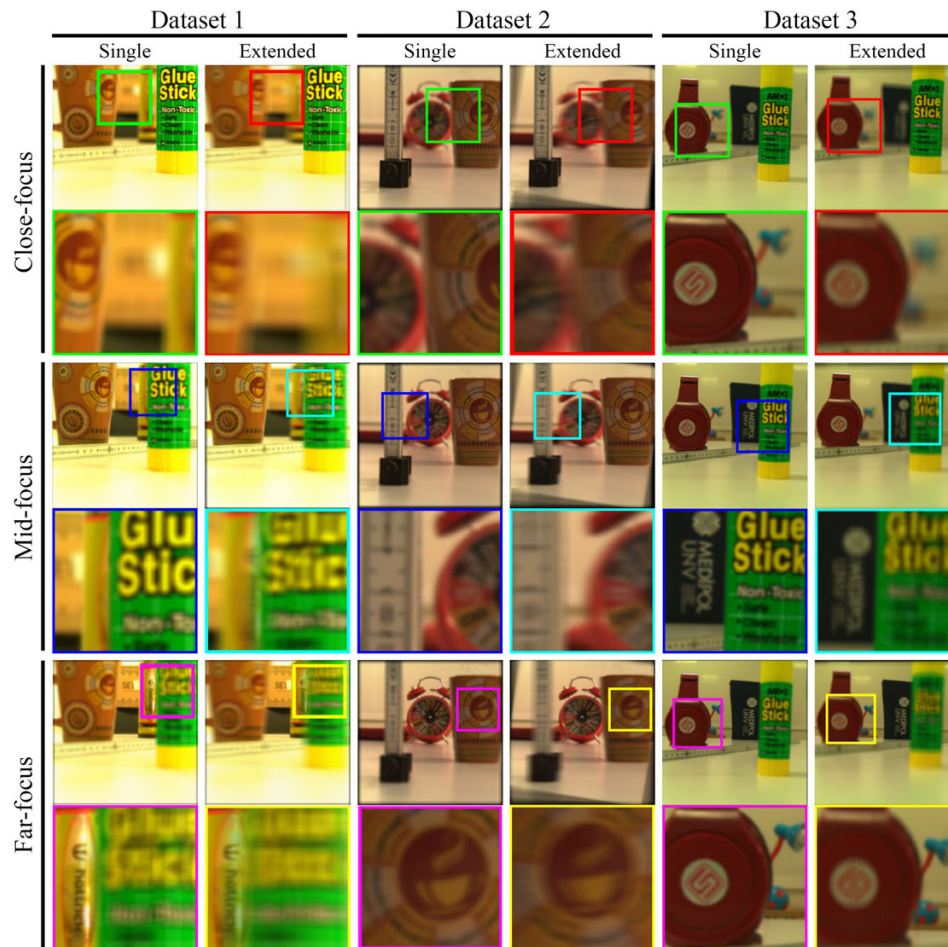


Figure 3.8: Comparison of refocusing and out-of-focus blurs at three different depths for single and extended light fields.

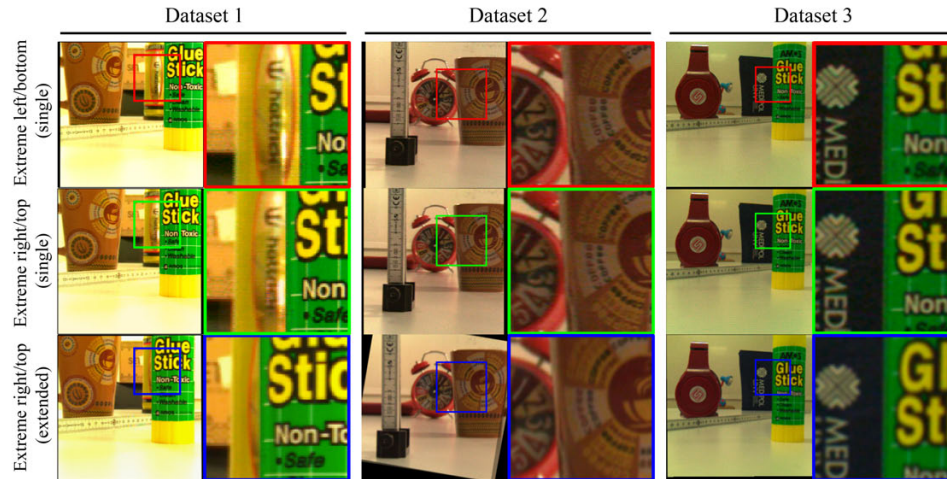


Figure 3.9: Translation parallax with the single light field and the extended light field. For first and second datasets: (Top) Leftmost sub-aperture image in the single light field and the extended light field; (Middle) Rightmost sub-aperture image in the single light field; (Bottom) Rightmost sub-aperture image in the extended light field. Whereas, for third dataset: (Top) Bottommost sub-aperture image in the single light field and the extended light field; (Middle) Topmost sub-aperture image in the single light field; (Bottom) Topmost sub-aperture image in the extended light field.

The relation between baseline and depth estimation accuracy for a stereo system has been studied in [39], where it is shown that the depth estimation error is inversely proportional with the baseline and increases quadratically with depth. By extending light field aperture, the baseline is essentially increased, which inherently improves both depth estimation range and accuracy. In Figure 3.10, the disparity map is shown, obtained by optical flow estimation technique [40] between the leftmost and rightmost sub-aperture images, for single and extended light fields. As seen in the figure, the range of the disparity map for the extended light field is (about three times) larger than that of the single light field.

3.5 Discussion and future work

In this chapter, a light field registration algorithm is presented to merge multiple light fields, obtaining extended synthetic aperture, which will yield an enhanced

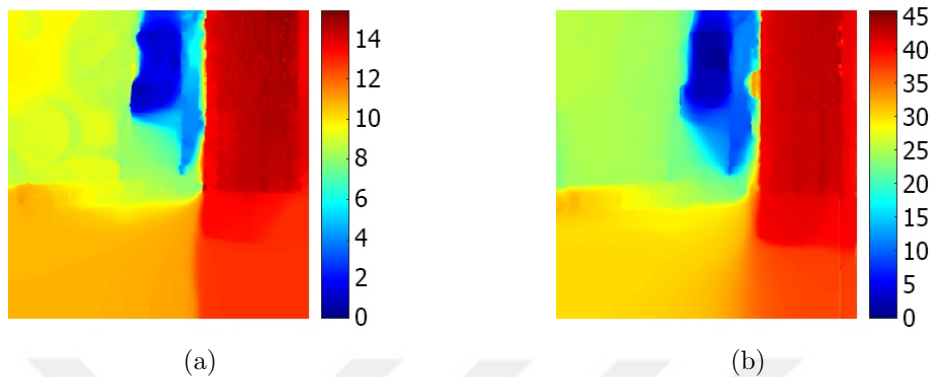


Figure 3.10: Disparity map comparison of single and extended light fields for Dataset 1.

signal-to-noise ratio. The method is tested with light field data captured by a Lytro camera, which makes the problem more challenging due to its low spatial resolution. One possible extension of the proposed method is increase angular resolution in addition to angular range. This can be done through defining a finer grid for interpolation. Another possible extension is to improve spatial resolution through interpolation in spatial domain in addition to interpolation in angular domain. We believe the proposed registration approach can be utilized in other applications, such as light field video compression and light field object tracking, as well.

Chapter 4

Hybrid-sensing for resolution enhancement

As discussed in the preliminary sections, the capabilities of a micro-lens array based light field cameras are limited by the fact that there is an essential trade-off between spatial and angular resolution. To achieve enough angular resolution, spatial resolution is compromised, resulting in visual quality that is far below today's standards. In this chapter, a hybrid sensing system is presented by using a regular high-resolution sensor in combination with a light field sensor which utilizes a single lens with minimal optics to achieve high spatial resolution light field. The use of a single lens prevents potential multi-lens problems, such as occlusions and artifacts due to mismatching lens aberrations. In the experiments, it is shown that the proposed hybrid-sensor camera leads to improved depth estimation in addition to increase in spatial resolution.

4.1 Related work

The issue of low spatial resolution in the context of light field gained significant attention among researchers. Various approaches have been proposed to address

this issue. Some of the approaches utilize prior knowledge of light field to reconstruct higher resolution perspectives [41], [42], [43], [44]. Few methods modify the hardware and play with the spatio-angular resolution trade-off to enhance spatial resolution [13], [12], [14]. There are also methods that use learning based techniques to improve the quality of light field perspectives [4], [45], [46].

Alternative to the such methods, there is a hybrid two-camera system proposed in [47], which includes a regular camera and a light field camera. The spatial resolution of the image captured with the regular camera is transferred to the sub-aperture images captured with the light field camera. In this approach, the spatial resolution of the light field can be increased to that of the regular camera. While the positions of light field and regular camera are arbitrary in [47], they are placed as a stereo system in [48], which enables faster processing through pre-calibration. Such a hybrid stereo system also allows extended and more accurate depth estimation capability due to larger baseline [48]. With the two-camera systems, the main problem is occlusion due to different viewpoints of the cameras. The problem can be alleviated by using a beam-splitter before the cameras to have the same viewpoints [49], [50], [45]. One challenge that multi-lens systems should address is the need to register and compensate for optical distortions when non-identical objective lenses are used.

The optical system that is presented in this paper utilizes a single lens to prevent potential problems, such as occlusion and geometric distortions, that multi-lens systems may have. The use of a single lens also enables a more compact camera design.

4.2 Optical design

The optical design presented in this research is composed of two image sensors (a regular sensor and a light field sensor), a 50:50 beam-splitter, and a 60mm converging lens. Light field sensor is extracted from a first generation Lytro camera with its embedded processor. A raw image downloaded from the light

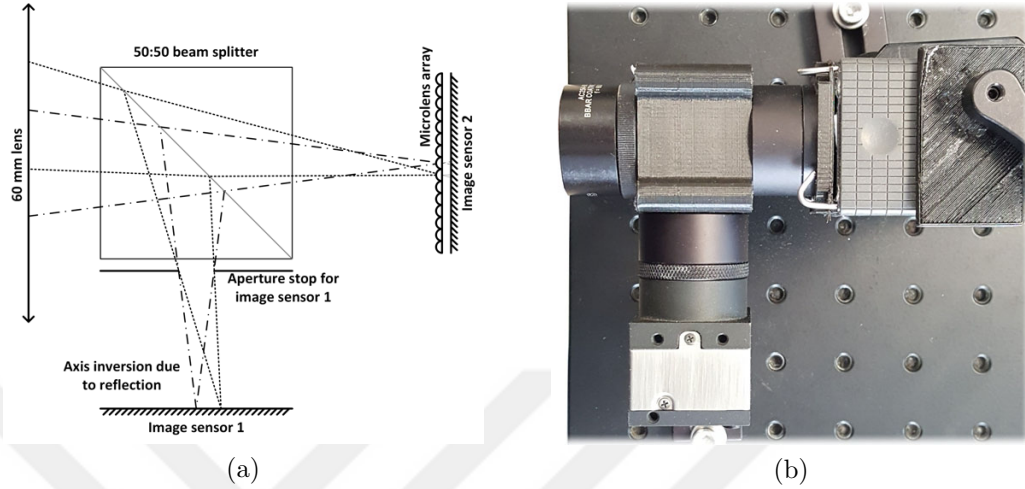


Figure 4.1: (a) Proposed optical design; (b) Top view of the hardware setup based on the optical design.

field camera is of resolution 3280×3280 pixels (10.7 megapixels). The captured image roughly consists of $328 \times 393.6 \approx 129,100$ micro-lenses, which becomes the spatial resolution after decoding the light field. The pixels beneath each micro-lens determines angular resolution, which roughly creates 9×9 perspectives. A CMOS sensor (DCC1240c) from Thorlabs [51] is utilized as a regular sensor, which gives a resolution of 1280×1024 pixels (1.3 megapixels).

A beam-splitter is used to equally divide the converging light rays on both sensors. As shown in Figure 4.1a, the image gets inverted when reflected in the the splitter. The position of regular sensor is adjusted to form a sharp image based on the object distance and focal length of the objective lens. The distance between the micro-lens array and the objective lens is also manually adjusted to have the imaged object within the depth-of-field of the light field camera.

Horizontal and vertical alignment of these sensors with respect to optical axis is necessary to get overlapping spatial regions in both sensors. The field of view of regular sensor is larger than that of light field sensor due to the size of their sensors. Hence cropping of the regular sensor image is employed to match its spatial range with the light field image.

To attain maximum angular resolution of the light field camera, the f-number

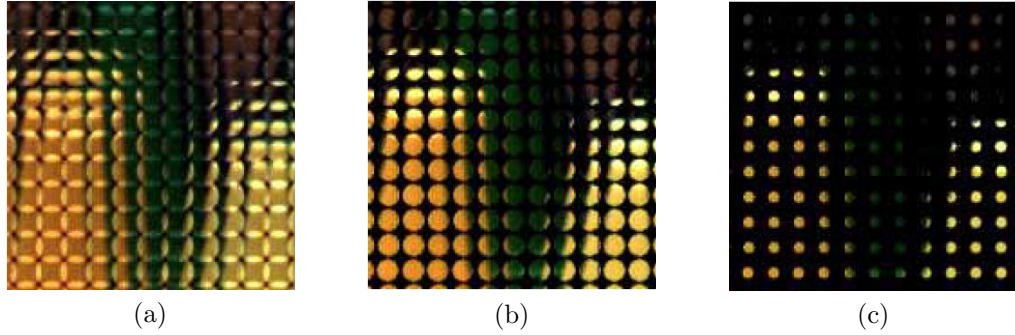


Figure 4.2: Close-up of micro-lens images formed with different main lens aperture sizes (a) $f/2.8$, (b) $f/4$ and (c) $f/8$. The micro-lenses are $f/4$.

of the objective lens should match with that of the micro-lenses [12]. When the f-number of the objective lens is more than that of the micro-lens, there is overlap between neighbouring micro-lens projections. When the f-number of the objective lens is less than that of the micro-lens, the micro-lens projections get highly vignetted, which results in loss of angular resolution. With the same f-numbers, the projections barely touch each other, which results in the maximum angular resolution. For our setup, the aperture size is limited by the diameter of the objective lens and it gives a 7×7 angular resolution which is good enough to demonstrate light field capabilities.

For ideal resolution transfer at a later stage, the imaging conditions for the regular sensor is made same with the middle perspective of the light field image, such that the middle perspective of the light field differs only in resolution with the regular sensor image. Since the depth of focus of the middle perspective is larger compared to the regular sensor, the aperture size of the regular sensor is reduced by placing a 4mm aperture stop just at the wall of beam-splitter, facing the regular sensor as shown in Figure 4.1.

4.3 Spatially enhanced light field creation

The proposed light field resolution enhancement algorithm is illustrated in Figure 4.3. This process is broken in three stages to achieve a high spatial resolution

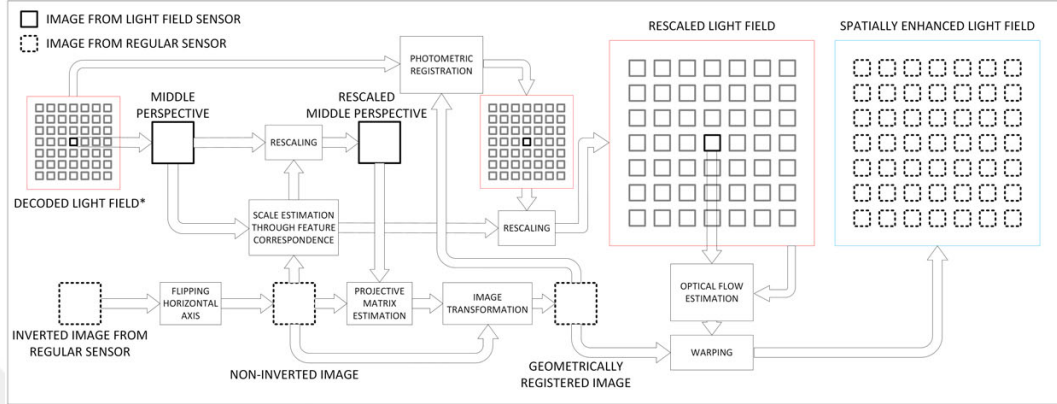


Figure 4.3: Illustration of the proposed light field resolution enhancement algorithm. Raw lenslet image is decoded using MATLAB toolbox [3] to create 4D plenoptic function.

light field, which are described in the following sub-sections.

4.3.1 Light field capture and pre-processing

The angular resolution of image captured using light field sensor is utilized in combination with the high spatial resolution of regular sensor to form a high-resolution light field. It requires one capture from each of the light field and regular sensors. Light field is captured with an exposure time of 1/8 seconds.

Due to misalignment between micro-lens array and sensor in a light field sensor a calibration step is mandatory before decoding. The calibration step moves the projection centers of micro-lenses on a regular integer grid with a corrected rotation. A lenslet image of a diffused white scene is captured by the light field sensor, which is required to computationally find out the actual centers of micro-lens projections, as described in [3]. For decoding, the MATLAB toolbox provided by Dansereau [3] is utilized. The decoding process involves registering the raw image on a regular grid and then slicing it to form a four-dimensional light field function for each color channel.

An increased exposure time is required for the regular sensor since the aperture size is reduced to achieve an all-in-focus image. Hence, the image from the regular

sensor is captured with an exposure time of 1.2 seconds. A downside of the long exposure time is that hot pixels may arise in the captured image. The image is horizontally flipped in software to compensate for the image inversion due to reflection by the beam-splitter.

4.3.2 Geometric registration

To achieve similar visual range in both the captures (from regular sensor and middle perspective from light field sensor), a homographic registration is done. It is required to keep the area of interest in the high-resolution image to its actual resolution. Before warping the high-resolution image to the middle perspective, the middle perspective is first scaled up to match the scale of the high-resolution image. The scale is estimated with the help of the *similarity* matrix, which is calculated by matching SURF features [52] in both the images.

The homographic transformation is estimated in both images to correct for planar transformations. The matching SURF features are extracted again in both images. The outliers within the set of matched features are removed by fitting them on the projective transformation model [32].

4.3.3 High-quality perspective formation

With a very small baseline in micro-lens array based light field camera, occlusions will be minimal as well. After resizing each of the light field perspective with the scale extracted in the previous section. The motion of each pixel in each perspective with respect to the middle perspective is estimated using the optical flow algorithm given in [40]. Before estimating motion vectors, the color distribution in each perspective is mapped to the middle perspective using a histogram based photometric registration algorithm [53]. For the sake of quality comparison, instead of mapping colors to the middle perspective, each perspective is mapped to the high-resolution image color space. To achieve a photometrically corrected



Figure 4.4: Photometric registration of light field images and regular sensor image. (a) Middle sub-aperture light field image before photometric registration; (b) Middle sub-aperture light field image after photometric registration; (c) Homographically corrected image from regular sensor taken as the reference for photometric registration.

image, the histogram of the image intensity is initially equalized using its inverse and then an inverse mapping of the target image intensities is applied to get a photometrically registered image. A result of such photometric mapping is shown in Figure 4.4.

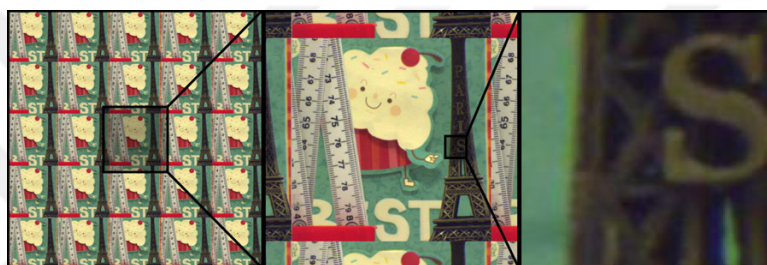
After motion vectors are estimated for each of the perspective, the high-resolution image is warped to create high quality perspectives. A comparison between a low-resolution light field and its high-resolution version is shown in Figure 4.5. The spatial resolution of the original light field is 380×380 pixels. The resulting spatial resolution of the light field is 1000×1000 pixels.

4.4 Experimental results

The enhanced resolution improves the performance in light field applications. One such application is depth estimation, which is also illustrated in this section along with resolution enhancement. The comparison in spatial resolution is shown between low and high-resolution light fields by refocusing at three distinct depths in two different datasets. (See Figures 4.6 and 4.7.) It can be clearly seen that sharpness of the refocused depths in high-resolution light field is relatively higher compared to the sharpness in low-resolution light field. To account for the better



(a)



(b)

Figure 4.5: Spatial resolution enhancement of light field. (a) Low-resolution bilinearly interpolated light field captured by the light field sensor; (b) Spatially enhanced light field as a result of the proposed algorithm.

qualitative analysis of resolution enhancement, the resolution enhancement is also demonstrated using a resolution chart as shown in Figure 4.8. In the resolution chart the converging lines can be better distinguished in the high-resolution image compared to the low-resolution image.

An increased sharpness is achieved because of the reduced pixel pitch in the higher resolution light field. The improvement in spatial resolution gives rise to more accurate depth estimation. A popular approach to estimate depth from light field data is the use of epipolar plane images (EPIs). A sharper image generally contains more features, therefore the lines formed by sharp features in the EPI are also sharp; and this eventually improves the depth estimation accuracy. The disparity map generated using a pair of low and high-resolution light fields for both datasets is shown in Figure 4.9. Visually comparing these maps will conclude the fact that the disparity map generated through high-resolution light field is more accurate than the one with low-resolution light field.

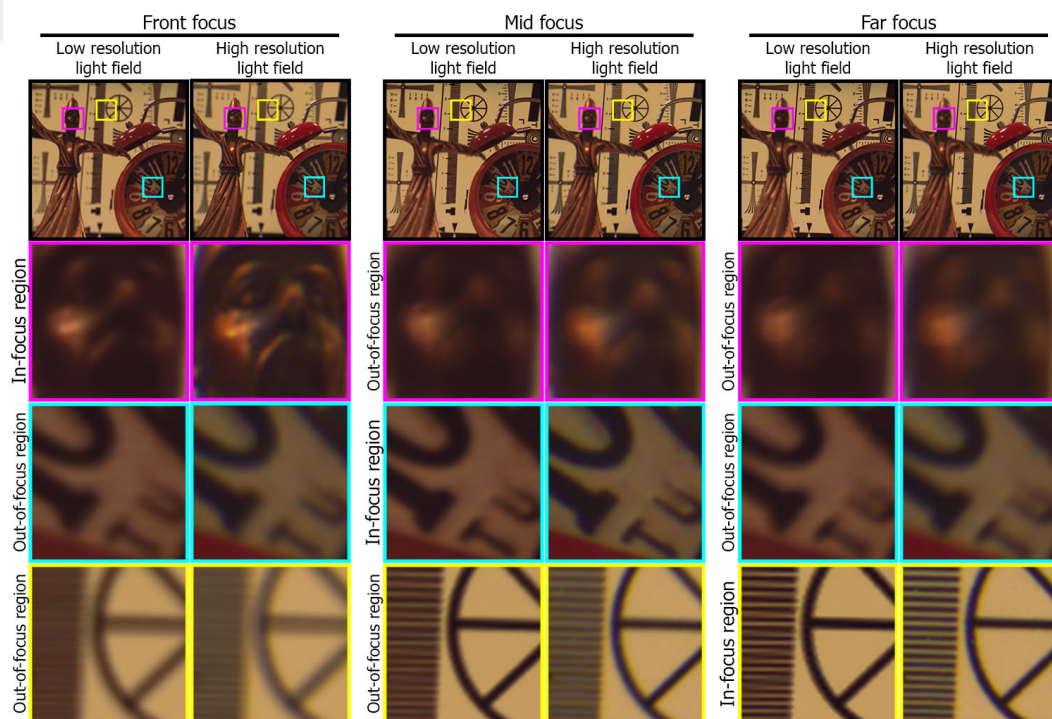


Figure 4.6: Spatial resolution comparison between original and spatially enhanced light fields for digital refocusing for Dataset 1. Using the shift-and-sum technique, light fields are focused at three different depths for both low-resolution and high-resolution light fields.

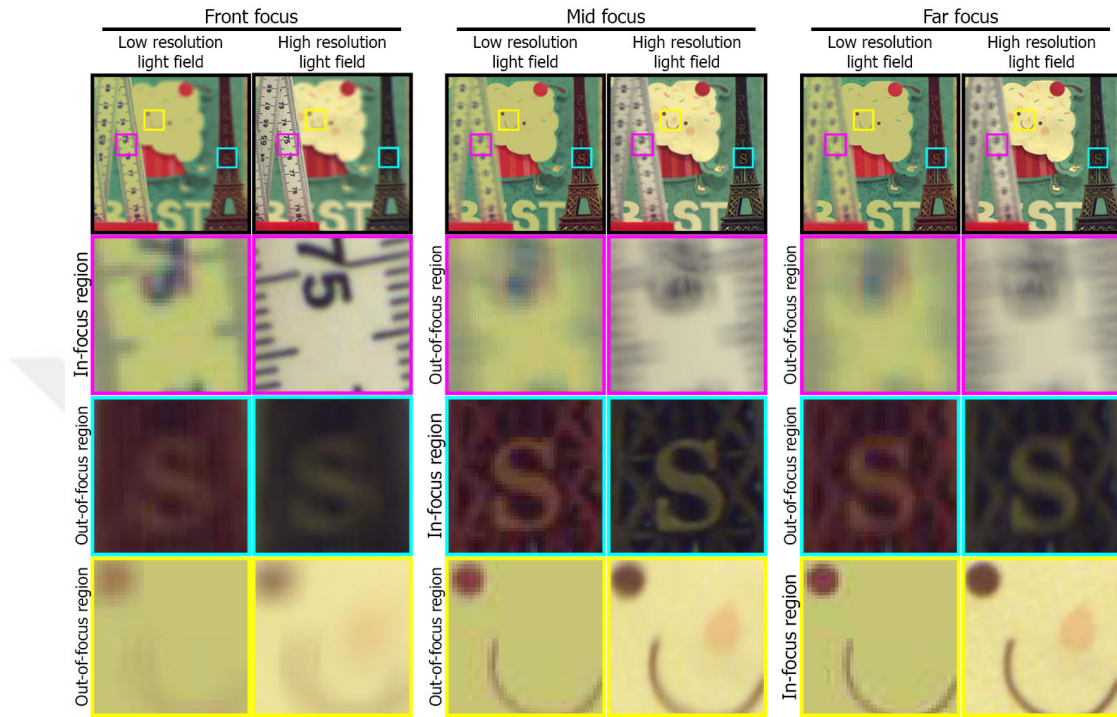


Figure 4.7: Spatial resolution comparison between original and spatially enhanced light fields for digital refocusing for Dataset 2. Using the shift-and-sum technique, light fields are focused at three different depths for both low-resolution and high-resolution light fields.

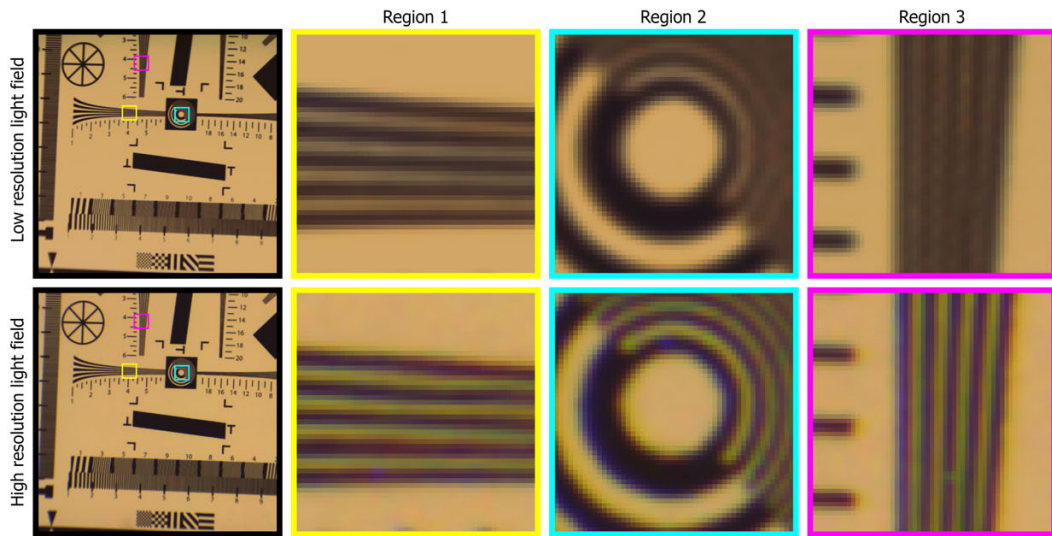
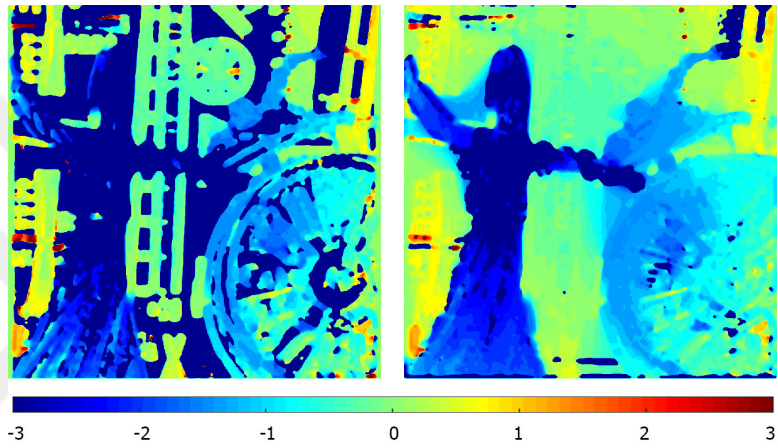
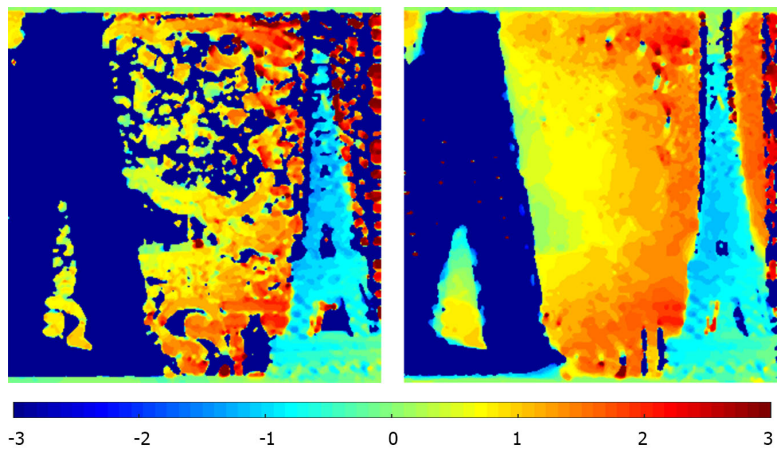


Figure 4.8: The quality of both light fields are compared after refocusing each of them to the current depth. At three different regions it is clearly visible that the lines are distinguishable pointing to high spatial resolution of reconstructed light field.



(a)



(b)

Figure 4.9: Quality of depth reconstruction is presented between (Left) low-resolution light field (Right) high-resolution light field; for two datasets.

4.5 Discussion and future work

In this research, a hybrid sensor light field imaging system is presented, which includes a regular and a light field sensor. An algorithm is presented to create high-resolution light field with the images captured by the system. A spatial resolution of 1000×1000 pixels is achieved, which is roughly seven times superior than the original light field. The resolution enhancement depends on numerous factors including the resolution of the regular sensor and ratios between the size of the sensors. Even with a very small baseline between the perspectives, translation parallax is present, which can create ghosting artifacts in some occluded regions.

Chapter 5

Controlled micro-shifts of light field sensor for resolution enhancement

A major problem of micro-lens array based light field cameras is low spatial resolution. As an alternative to the hybrid sensor approach proposed in the previous chapter, we introduce another approach in this chapter to overcome the low spatial resolution problem of light fields captured with micro-lens array based cameras. Traditional super-resolution techniques use multiple image captures and registration on a finer resolution grid to achieve higher spatial resolution. In the microscanning approach, the image sensor is shifted by specific amounts before each capture to ensure the necessary spatial diversity among the input images. In this chapter, the same idea is used in the context of light field resolution enhancement. The micro-lens array structure of the light field sensor requires specific shift pattern design, which is discussed in the chapter. Also, the relation between the number of light field captures and the resolution enhancement factor is analyzed.

5.1 Related work

Various techniques have been proposed to computationally tackle the light field resolution enhancement problem. In [41], texture preserving priors are used to generate high-resolution perspectives, with the assumption that scene contains Lambertian surfaces only. In [44], Levin *et al.* utilize frequency domain interpolation to enhance resolution without using depth information. In [42], high-resolution perspectives are created under the assumption of continuous linear epipolar lines. In [43], Mitra *et al.* employ Gaussian Mixture Model (GMM) learning for light field patches and obtain super-resolved light field through Bayesian estimation.

Artificial neural networks are also utilized in some of the implementations for both angular and spatial resolution enhancement. In [4] and [45], dictionary based learning is employed. In [46], Yoon *et al.* present a deep convolution neural network based implementation for enhancing both spatial and angular information up to a factor of two.

Light field and conventional cameras together have also been utilized to enhance the low-resolution light field perspectives using the high-resolution information from the conventional camera. In [47], Boominathan *et al.* utilize patch based registration, while Alam *et al.* utilize optical flow based registration in [48]. The optical designs of [49] and [50] are similar, in which these two cameras are utilized together by splitting light rays coming to the cameras using a beam-splitter. In [49], Wang *et al.* generate high-resolution light field with the help of complex steerable pyramid filters and phase modification.

Alternative camera designs have also been proposed to generate high-resolution perspectives. In [13], Georgiev *et al.* present “Plenoptic 2.0” camera, whose main idea is to increase the spatial resolution by placing the micro-lens array at a different depth in comparison with the original light field camera design [12]. In contrast to [12], the Raytrix camera [14] uses a micro-lens array that is composed of different focal lengths to increase the depth of field.

In our method, multiple light fields are captured, where a micro-shift is introduced on the in-plane position of the light field sensor before each capture, while keeping the other optical elements static. These micro-shifts are produced with the help of piezo-electric micro-actuators. As a result, the spatial resolution of the light field is enhanced without relying on complex algorithmic implementations.

5.2 Enhancing spatial resolution through micro-shifted light field sensor

5.2.1 The microscanning idea

The light field spatial resolution for a micro-lens array based light field camera [12] is determined by the number of micro-lenses in the micro-lens array. In a standard Lytro camera, there are approximately 0.1 million micro-lenses packed in a hexagonal grid in front of an image sensor. It is possible to increase the number of micro-lenses; however, that would result in a decrease in angular resolution [13].

In this work, the light rays are recorded at a finer grid through multiple captures and by shifting the light field sensor before each capture. The idea is known as microscanning, and it has been used for spatial resolution enhancement of traditional cameras. Figure 5.1 illustrates the use of this idea for light field resolution enhancement. In the illustration, two light fields are captured with a controlled vertical shift of micro-lens array. Here, the shift is chosen to be the half of the distance between the centers of adjacent micro-lenses. The combination of these two light field captures forms a light field having more micro-lenses with a finer pitch, resulting in a spatial resolution twice as much as the previous one. The idea can be extended to two dimensional case and more samples can be taken to improve the spatial resolution.

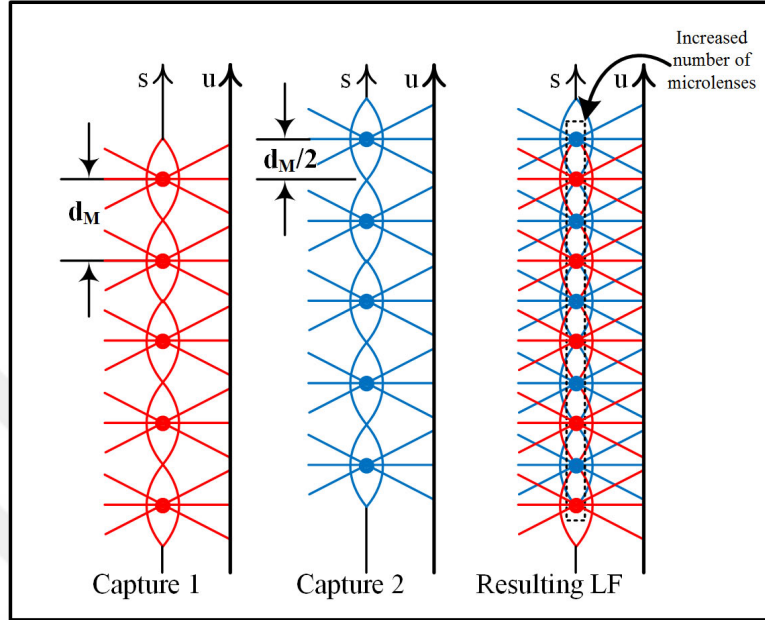


Figure 5.1: Spatial resolution improvement of light field due to increased number of micro-lenses by merging two light field captures with a shift equal to the half of d_M (the distance between the centers of adjacent micro-lenses).

5.2.2 Optical design and light field acquisition

The idea suggests that a detached light field sensor (combination of a micro-lens array and a photosensor separated by the focal length of the micro-lenses) is required for shifting it during captures. In our experiments, we used a light field sensor retrieved from a first generation Lytro camera. The sensor within the light field sensor has an effective resolution of $3,280 \times 3,280$ pixels [2]. This optical design is shown in Figure 5.2.

To create the micro-shifts in horizontal and vertical directions, two Thorlabs motorized translation stages are used. The product datasheet claims an accuracy of $2.0 \mu\text{m}$ [51]. The product comes with an OEM software to control it. Our setup is composed of a 60mm achromatic lens (as the objective lens) with an aperture to match the f-number of the micro-lenses, and the light field sensor that is fixed to the translation stage.

The micro-lenses on the Lytro light field sensor are arranged as a hexagonal

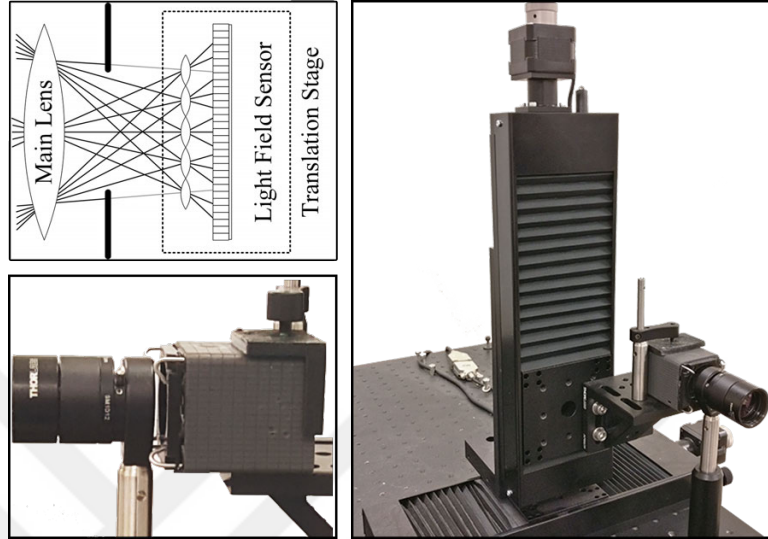


Figure 5.2: (Top-left) Proposed optical design; (Bottom-left) Optical hardware (zoomed-in side view); (Right) Optical hardware showing the translation stages.

grid, as shown in Figure 5.3. There is a region of 9×9 pixels behind each micro-lens which forms 9×9 perspectives. The distance between the centers of each adjacent micro-lens is $14 \mu\text{m}$; the vertical distance between adjacent two rows of micro-lens array is $12.12 \mu\text{m}$. The hexagonal arrangement of micro-lenses produces hexagonally gridded pixels for each perspective. The shifts are designed to have equally spaced pixels in a rectangular grid in the resulting light field. As listed in Table 5.1, 16 raw light fields are captured with controlled shifts equal to the integer multiple of $3.5 \mu\text{m}$ in horizontal direction and of $3.03 \mu\text{m}$ in vertical direction. Ideally, with this combination, the spatial resolution gets 16 times enhanced as shown in Figure 5.4e. For later comparison, four sets of light fields are generated by utilizing different subsets of 16 captures. These combinations are generated by picking two, four, eight and sixteen light fields from the set of 16 captures as shown in Figures 5.4b, 5.4c, 5.4d and 5.4e respectively.

Table 5.1: Translation amounts required to build a 4 times enhanced spatial grid in each direction, analogous to Figure 5.4e, with 16 light field captures.

Cap.	$T_u(\mu m)$	$T_v(\mu m)$	Cap.	$T_u(\mu m)$	$T_v(\mu m)$
1	0.00	0.00	9	3.50	-21.21
2	0.00	-3.03	10	3.50	-18.18
3	0.00	-6.06	11	3.50	-15.15
4	0.00	-9.09	12	3.50	-12.12
5	0.00	-12.12	13	3.50	-9.09
6	0.00	-15.15	14	3.50	-6.06
7	0.00	-18.18	15	3.50	-3.03
8	0.00	-21.21	16	3.50	0.00

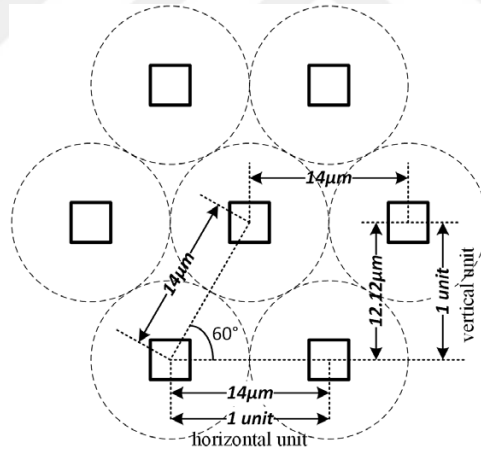


Figure 5.3: Calibrated lenslet image with squares representing center pixel of each micro-lens (dotted circle). (Not drawn to scale.)

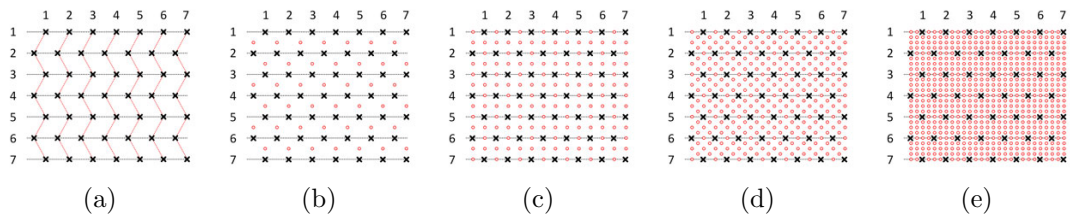


Figure 5.4: Combination of translated images to form high-resolution perspectives with the number of light field equal to: (a) 1; (b) 2; (c) 4; (d) 8; (e) 16.

5.3 Reconstructing high-resolution light field

5.3.1 Calibrating the lenslet grid

The raw light field capture (lenslet image) needs to be calibrated to properly decode a 4D light field. There may exist misalignment between the micro-lens array and the image sensor. This effect is visible in the projection of micro-lenses on the lenslet image. For decoding, the center of each projected micro-lens must be known. A better approach is to align the micro-lens centers on an integer grid, which is achieved by transforming the lenslet image. micro-lens centers can be efficiently determined through capturing a white scene, and then, determining the local brightest point to represent the micro-lens center. This entire process is explained in [3]. The rotation, offset and lenslet spacing of the lenslet image is estimated with the help of white image; the estimated parameters are then utilized to decode each light field capture with the same setup.

5.3.2 Light field registration

If the shifts produced by the actuators were perfect, then the pixel values from the captured light fields would be placed at their proper locations. However, in practice, the shifts are not perfectly accurate; and we expect artifacts, especially in texture and edge regions, if the light fields are not registered properly. The problem can be addressed by estimating the shifts between the input light fields before fusing the light fields.

The problem of image registration has been studied extensively over the years. In case of estimating only the translation between the images, the fast Fourier transform (FFT) based algorithm provided in [5] can be used. It obtains an initial estimate of the cross-correlation peak using FFT and then refines the estimates through upsampling the Fourier transforms.

As we will see later in the chapter, there may still exist some minor artifacts

even after the registration process. To reduce these artifacts, each perspective image is separately formed for even and odd rows of the micro-lenses in the lenslet image; these perspectives are then averaged to reduce the artifacts.

5.3.3 Non-uniform interpolation

The non-uniform grid interpolation is performed to interpolate on a regular grid using densely distributed scattered pixels. To interpolate using densely distributed pixels, a finer resolution grid (1512×1296) is used, compared to the initial resolution of 378×328 pixels. For this purpose, the method presented in [54] is employed, which is based on interpolating intensities in regular grid using Delaunay triangulation, to create high-resolution perspectives of light field.

After interpolation, high-resolution perspectives are achieved whose quality depends on the number of light fields used to create them. If the number of input light fields is reduced, the triangles formed using Delaunay triangulation become larger. And larger triangles suggest poor estimate of the actual value due to unavailability of the nearby pixels. At the end, 7×7 high-resolution perspectives are created.

5.3.4 Blind deblurring

During light field acquisition, a bundle of rays passes through a portion of the main lens, and then gathered and focused by a micro-lens on a pixel. In other words, the effective pixel size is the area of micro-lens. A portion from this bundle is also gathered by the micro-lens when shifted for the next capture. Since the micro-lens plane represents the spatial resolution and there is overlap in the micro-lens plane between shifted captures, there exists a spatial blur in the resulting light field. To reduce the effect of this blur, the high-resolution perspectives are blindly deblurred using the MATLAB function “deconvblind”. As an initial estimate of blur, we use a Gaussian kernel with size 5×5 pixels and

a standard deviation of 1.

5.4 Experimental results

In this section, we demonstrate the spatial resolution enhancement with qualitative and quantitative results. We investigate the effect of the number of input light fields on resolution enhancement, and compare with basic light field decoding techniques presented in [3] and [4]. Refocusing results are also presented between bicubically resized and resolution enhanced light fields. The quantitative analysis is based on peak signal to noise ration (PSNR). At the end, the effect of using inaccurate shift in the formation of high-resolution perspective is presented.

5.4.1 Qualitative analysis

5.4.1.1 Producing finer light field grid

Through the application of controlled shifts of light field sensor, more micro-lenses are effectively placed within the same range of the image sensor. Figure 5.5a shows the lenslet image of a single light field, which has 378×328 micro-lenses, whereas in Figure 5.5b, the lenslet image of the spatially enhanced light field has 1512×1312 micro-lenses within the same image sensor range.

5.4.1.2 Comparison with base techniques

Figure 5.6 demonstrates that the quality of the light field (in terms of spatial resolution) generated by merging 16 light fields is much higher when compared with basic decoding implementations of [4] and [3].

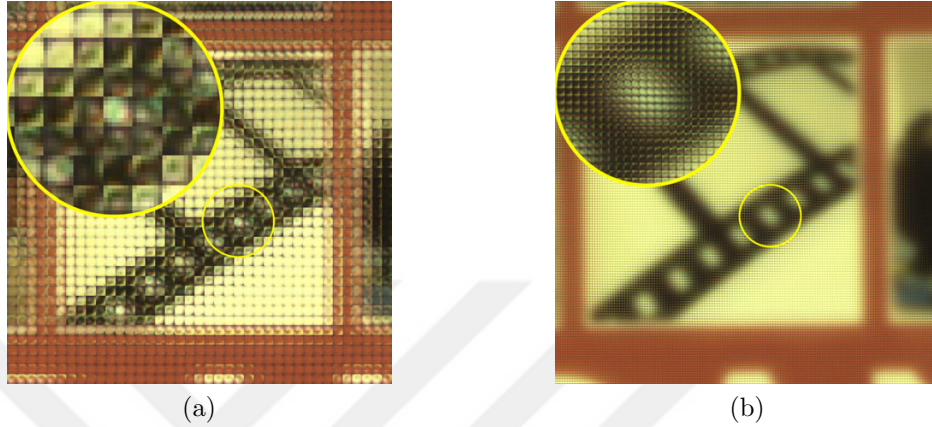


Figure 5.5: Lenslet images with their magnified patches illustrating micro-lens distribution. (a) Low spatial resolution light field; (b) High spatial resolution light field by merging 16 light field captures.

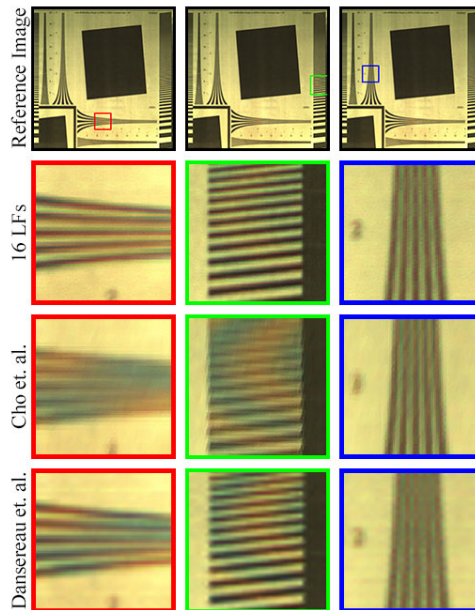


Figure 5.6: Visual comparison of quality of middle perspectives between light fields generated with implementations of Dansereau *et al.* [3] and Cho *et al.* [4].

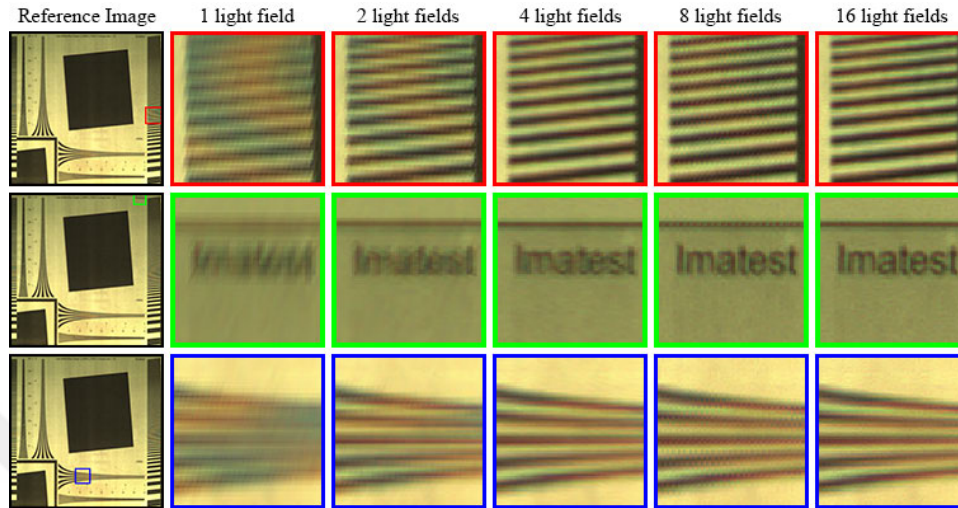


Figure 5.7: Visual comparison of quality of middle perspectives between light fields generated with different number of captures.

5.4.1.3 Effect of increasing the number of light fields

During the experiment, 16 light fields are captured with the controlled shifts of light field sensor as described in Table 5.1. Five light fields are generated by picking five different subsets of captured light fields as shown in Figure 5.4. These light fields are created by merging 1, 2, 4, 8 and 16 light fields. In Figure 5.7, a visual comparison is provided to highlight the change of spatial quality as the number of light fields are increased. From this figure, it can be visualized that sharpness of resulting light field in spatial domain is proportional to the number of light fields utilized to generate it.

5.4.1.4 Post-capture refocusing

The basic application of post-capture refocusing of light field technology comes into play when a real-world scene containing multiple depths is available. Post-capture refocusing is demonstrated in Figure 5.8, which features the comparison between light field generated with 16 captures, a bicubically enhanced light field and the light fields generated through the basic decoding techniques [3] and [4].

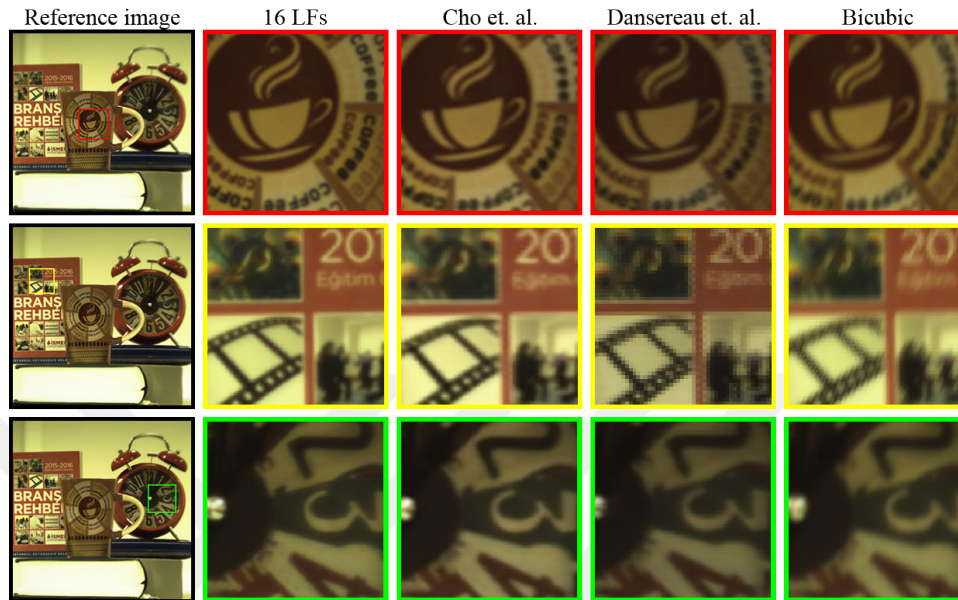


Figure 5.8: Comparison of post-capture refocusing demonstrated for light fields generated through (1) merging 16 captures, at three different depths, (2) Cho *et al.* [4] implementation, (3) Dansereau *et al.* [3] implementation, and (4) bicubic resizing.

5.4.2 Quantitative comparison

Figure 5.9 provides a PSNR comparison for different number of light field captures as well as the base light field decoding techniques. It is expectedly found that the spatial resolution of light field created through 16 captures is superior from the light fields obtained with less number of captures.

5.4.3 Computation time

The implementations are done in MATLAB on a PC with an Intel Core i5-4570 processor clocked at 3.20 GHz with 12 GB RAM. Figure 5.10 illustrates that, for same spatial resolution, the computational time required for interpolation is monotonically increasing for each perspective image as the number of light fields used is increased.

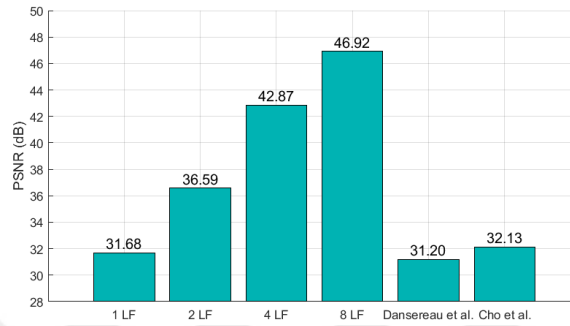


Figure 5.9: Comparison of quality between high-resolution perspectives created with different number of captures based on probabilistic signal-to-noise ratio is presented by considering high-resolution perspective generated with 16 captures as a reference.

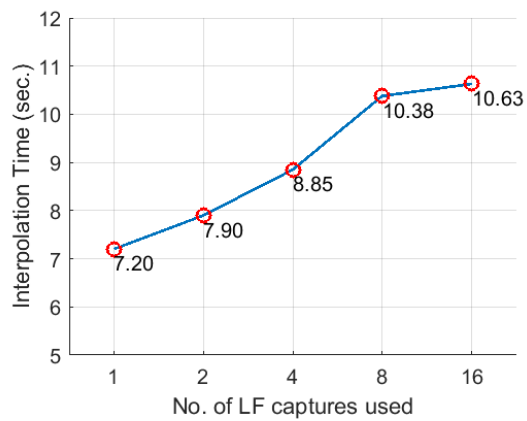


Figure 5.10: The time required to generate single high-resolution perspective for interpolation using different number of captures.

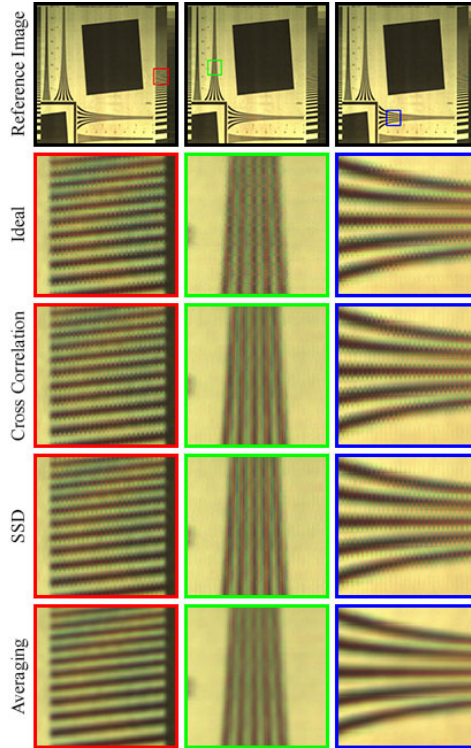


Figure 5.11: Comparison of different techniques to eliminate zig-zag artifacts due to inaccurate shifts. High-resolution perspectives are reconstructed using 16 light field captures, where: (Ideal) Actual shifts are applied with the assumption of zero translation error; (Cross Correlation) Shifts estimated using [5]; (SSD) Shifts estimated by minimizing sum-of-squared difference (SSD) between the images; (Averaging) By separately forming and averaging perspectives from even and odd rows of the micro-lens images, where the shifts are estimated using the SSD technique.

5.4.4 Effect of inaccurate shifts

High frequency zig-zag artifacts can be seen in Figure 5.11, when it is assumed that the micro-actuator shifts are accurate, and there is no registration process applied. These artifacts can be reduced using cross correlation or sum-of-squared differences (SSD) based registrations. By separately forming and averaging perspectives formed from even and odd rows of the micro-lens images, these artifacts can be further reduced.

5.5 Discussion and future work

In this chapter, a technique to enhance spatial resolution through controlled micro-shifts of light field sensor is presented. The optical setup and steps of the reconstruction algorithm are detailed, including calibration, registration, interpolation and reconstruction to create high-resolution perspectives. The high frequency artifacts due to inaccurate shifts are reduced through averaging; in future, the registration process can be revised to reduce the artifacts even further.

Chapter 6

Conclusions

A first generation Lytro camera is used in this thesis. Its architecture utilizes micro-lens array based light field camera design. Despite being compact and low-cost solution, it has de-facto problems of limited aperture size and spatial resolution. And in this thesis, we have presented some techniques to overcome the mentioned limitations.

The compact form factor of a light field camera limits the maximum achievable aperture size. It is known that the accuracy of depth estimation is proportional to the size of the aperture. Therefore, the aperture size poses a significant limitation on the quality of estimated depth. In the third chapter, we presented a technique for synthetically extending aperture size to cope with this problem. The extension of the aperture is achieved by stitching multiple light fields, having arbitrary camera motion during each capture. Before stitching, each perspective is registered to a co-planar surface by estimating correct orientation concerning a reference perspective. Later in the chapter, it is shown that by achieving extended aperture, the depth estimation accuracy is also improved.

The light field of the first generation Lytro camera has low spatial resolution. Apart from the low visual quality, this also degrades several machine vision applications, such as depth estimation. Therefore, two different approaches are

presented for spatial resolution enhancement. In the first technique, a combination of a light field sensor and a conventional sensor is used to create high spatial resolution light field through a motion estimator. The proposed optical design utilizes a minimum number of elements to avoid complex optical distortions. In the second technique, a microscanning approach is used in the context of light field. Each new capture is taken by applying some fractional shift to the position of micro-lens array to record light field information at intermediate locations. Thereby, the captured light fields are stitched together to form a high spatial resolution light field for both the techniques. Experimental results are provided to demonstrate the effect of resolution improvement.

Light field technology is relatively new, and there are still open problems to be addressed. The techniques presented in this thesis aim to relieve some of these problems. As a future work, the techniques can be extended through, for instance including dense optical flow algorithms, to handle dynamic scenes.

Bibliography

- [1] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” in *ACM Transactions on Graphics (TOG)*, vol. 24, pp. 765–776, ACM, 2005.
- [2] “Lytro, Inc.” <https://www.lytro.com/>.
- [3] D. G. Dansereau, O. Pizarro, and S. B. Williams, “Decoding, calibration and rectification for lenselet-based plenoptic cameras,” in *Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1027–1034, IEEE, 2013.
- [4] D. Cho, M. Lee, S. Kim, and Y. W. Tai, “Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction,” in *Int. Conf. on Computer Vision*, pp. 3280–3287, 2013.
- [5] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, “Efficient subpixel image registration algorithms,” *Transactions of the OSA on Optics Letters*, vol. 33, no. 2, pp. 156–158, 2008.
- [6] G. Lippmann, “Epreuves reversibles donnant la sensation du relief,” *Journal de Physique Thorique et Applique*, vol. 7, no. 1, pp. 821–825, 1908.
- [7] A. Gershun, “The light field,” *Journal of Mathematics and Physics*, vol. 18, no. 1, pp. 51–151, 1939.
- [8] E. H. Adelson and J. R. Bergen, *The plenoptic function and the elements of early vision*. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

- [9] M. Levoy and P. Hanrahan, “Light field rendering,” in *Int. Conf. on Computer Graphics and Interactive Techniques*, pp. 31–42, ACM, 1996.
- [10] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *Int. Conf. on Computer Graphics and Interactive Techniques*, pp. 43–54, ACM, 1996.
- [11] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, “A real-time distributed light field camera,” in *Eurographics Workshop on Rendering*, pp. 77–86, 2002.
- [12] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Computer Science Technical Report CSTR*, vol. 2, no. 11, pp. 1–11, 2005.
- [13] A. Lumsdaine and T. Georgiev, “The focused plenoptic camera,” in *IEEE Int. Conf. on Computational Photography*, pp. 1–8, 2009.
- [14] C. Perwass and L. Wietzke, “Single lens 3d-camera with extended depth-of-field,” in *Human Vision and Electronic Imaging*, vol. 17, 2012.
- [15] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, “Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing,” *ACM Trans. on Graphics*, pp. 1–12, 2007.
- [16] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. K. Nayar, and C. Intwala, “Spatio-angular resolution tradeoffs in integral photography,” in *Eurographics Conf. on Rendering Techniques*, pp. 263–272, 2006.
- [17] J. Unger, A. Wenger, T. Hawkins, A. Gardner, and P. Debevec, “Capturing and rendering with incident light fields,” in *Eurographics Workshop on Rendering*, pp. 141–149, 2003.
- [18] A. Manakov, J. F. Restrepo, O. Klehm, R. Hegedus, E. Eisemann, H. P. Seidel, and I. Ihrke, “A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging,” *ACM Trans. on Graphics*, vol. 32, no. 4, pp. 1–14, 2013.

- [19] M. U. Mukati and B. K. Gunturk, “Light field registration: A multi-view geometry approach,” in *Signal Processing and Communication Application Conference (SIU), 2016 24th*, pp. 1289–1292, IEEE, 2016.
- [20] M. U. Mukati and B. K. Gunturk, “Hybrid-sensor high-resolution light field imaging,” in *Signal Processing and Communication Application Conference (SIU), 2017 25th*, pp. 1–4, 2017.
- [21] L. Stroebel, J. Compton, I. Current, and R. Zakia, “Photographic materials and processes, 1986.”
- [22] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, “Using plane+ parallax for calibrating dense camera arrays,” in *Proceedings of IEEE in Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. I–I, IEEE, 2004.
- [23] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, “Synthetic aperture confocal imaging,” in *ACM Transactions on Graphics (ToG)*, vol. 23, pp. 825–834, ACM, 2004.
- [24] C. Birklbauer and O. Bimber, “Panorama light-field imaging,” in *Computer Graphics Forum*, vol. 33, pp. 43–52, Wiley Online Library, 2014.
- [25] X. Guo, Z. Yu, S. B. Kang, H. Lin, and J. Yu, “Enhancing light fields through ray-space stitching,” *IEEE Trans. on Visualization and Computer Graphics*, pp. 1852–1861, 2015.
- [26] O. Johannsen, A. Sulc, and B. Goldluecke, “On linear structure from motion for light field cameras,” in *Int. Conf. on Computer Vision*, pp. 720–728, 2015.
- [27] H. Li, R. Hartley, and J. H. Kim, “A linear approach to motion estimation using generalized camera models,” in *Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [28] N. Sabater, V. Drazic, M. Seifi, G. Sandri, and P. Perez, “Light-field demultiplexing and disparity estimation,” in *Tech. Report HAL-00925652*, 2014.
- [29] M. D. Grossberg and S. K. Nayar, “Determining the camera response from images: what is knowable?,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1455–1467, 2003.

- [30] R. C. Bolles, H. H. Baker, and D. H. Marimont, “Epipolar-plane image analysis: An approach to determining structure from motion,” *International Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.
- [31] R. O. Duda and P. E. Hart, “Use of the hough transformation to detect lines and curves in pictures,” *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.
- [32] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [33] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Alvey Vision Conference*, vol. 15, p. 50, Citeseer, 1988.
- [34] C. Tomasi and T. Kanade, *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991.
- [35] B. K. P. Horn, “Recovering baseline and orientation from essential matrix,” *Journal of the Optical Society America*, vol. 110, 1990.
- [36] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An invitation to 3-D vision*. Springer, 2004.
- [37] P. J. Rousseeuw, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.
- [38] E. Hille, *Analytic function theory*, vol. 2. American Mathematical Soc., 2005.
- [39] D. Gallup, J. M. Frahm, P. Mordohai, and M. Pollefeys, “Variable baseline/resolution stereo,” in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- [40] C. Liu, J. Yuen, and A. Torralba, “Sift flow: Dense correspondence across scenes and its applications,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, 2011.
- [41] T. E. Bishop, S. Zanetti, and P. Favaro, “Light field superresolution,” in *Proc. IEEE Int. Conf. Computational Photography*, pp. 1–9, 2009.

- [42] S. Wanner and B. Goldluecke, “Variational light field analysis for disparity estimation and super-resolution,” *Transactions of the IEEE on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.
- [43] K. Mitra and A. Veeraraghavan, “Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch prior,” in *Proceedings of the IEEE Computer Vision and Pattern Recognition Workshops*, pp. 22–28, IEEE, 2012.
- [44] A. Levin and F. Durand, “Linear view synthesis using a dimensionality gap light field prior,” in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pp. 1831–1838, IEEE, 2010.
- [45] J. Wu, H. Wang, X. Wang, and Y. Zhang, “A novel light field super-resolution framework based on hybrid imaging system,” in *Proceedings of the IEEE Visual Communications and Image Processing*, pp. 1–4, IEEE, 2015.
- [46] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. So Kweon, “Learning a deep convolutional network for light-field image super-resolution,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 24–32, 2015.
- [47] V. Boominathan, K. Mitra, and A. Veeraraghavan, “Improving resolution and depth-of-field of light field cameras using a hybrid imaging system,” in *Proceedings of the IEEE International Conference on Computational Photography*, pp. 1–10, IEEE, 2014.
- [48] M. Z. Alam and B. K. Gunturk, “Hybrid stereo imaging including a light field and a regular camera,” in *Proceedings of the IEEE Signal Processing and Communication Application Conference*, pp. 1293–1296, IEEE, 2016.
- [49] X. Wang, L. Li, and G. Hou, “High-resolution light field reconstruction using a hybrid imaging system,” *Transaction of the OSA on Applied Optics*, vol. 55, no. 10, pp. 2580–2593, 2016.

- [50] F. Dai, J. Lu, Y. Ma, and Y. Zhang, “High-resolution light field cameras based on a hybrid imaging system,” in *Proceedings of the ISOP SPIE/COS Photonics Asia*, pp. 92730K–92730K, ISOP, 2014.
- [51] “Thorlabs, Inc.” <https://www.thorlabs.com/>.
- [52] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [53] I. Hossain and B. K. Gunturk, “High dynamic range imaging of non-static scenes,” in *Proc. SPIE Electronic Imaging*, pp. 78760P1–8, 2011.
- [54] S. Lertrattanapanich and N. K. Bose, “High resolution image formation from low resolution frames using delaunay triangulation,” *Transactions of the IEEE on Image Processing*, vol. 11, no. 12, pp. 1427–1441, 2002.

EXTENDING LIGHT FIELD CAMERA CAPABILITIES

ORIGINALITY REPORT

7%

SIMILARITY INDEX

4%

INTERNET SOURCES

6%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1

Dai, Qionghai, Tsutomu Shimura, Feng Dai, Jing Lu, Yike Ma, and Yongdong Zhang. "High-resolution light field cameras based on a hybrid imaging system", Optoelectronic Imaging and Multimedia Technology III, 2014.

Publication

<1%

2

Zhang, Chi, Guangqi Hou, Zhenan Sun, and Tieniu Tan. "Efficient auto-refocusing of iris images for light-field cameras", IEEE International Joint Conference on Biometrics, 2014.

Publication

<1%

3

dobing.info

Internet Source

<1%

4

Xiang Wang, Lin Li, GuangQi Hou. "High-resolution light field reconstruction using a hybrid imaging system", Applied Optics, 2016

Publication

<1%

5

docslide.us

Internet Source

<1%
