

**ABSOLUTE DEPTH  
USING  
LOW-COST LIGHT FIELD  
CAMERAS**

By  
**Shreedhar Rangappa**

A Doctoral Thesis submitted in partial fulfilment of the requirements for the award  
of Doctor of Philosophy of Loughborough University

**January 2018**

## Acknowledgments

I would firstly like to convey my sincere thanks to my first supervisor, Dr Jon Petzing, for his guidance, support, patience and encouragement throughout various stages of my PhD journey. *I always imagined PhD journey as unseen track of a rally, being driver myself and Jon as my co-driver, who constantly guided me to stay on track and reach the goal.*

I wish to thank my second supervisor, Dr Peter Kinnell, for his wonderful ideas that really made me think of different perspectives of solving problems. Also, I wish to thank Intelligent Automation Centre and the Loughborough University for supporting throughout my PhD journey.

It would have been difficult to manage my thoughts, ideas and feelings without someone to listen and respond to it, which was extremely well handled by my friends and loved ones, to whom I convey my thanks.

Finally, thanks to my parents (RJ) for their unconditional love and  $\infty$  support.

ತಾಯಿ ಮತ್ತು ತಂದೆಗೆ ಅರ್ಪಿಸುತ್ತೇನೆ



Dedicated to mom & dad

## **Abstract**

Digital cameras are increasingly used for measurement tasks within engineering scenarios, often being part of metrology platforms. Existing cameras are well equipped to provide 2D information about the fields of view (FOV) they observe, the objects within the FOV, and the accompanying environments. But for some applications these 2D results are not sufficient, specifically applications that require Z dimensional data (depth data) along with the X and Y dimensional data. New designs of camera systems have previously been developed by integrating multiple cameras to provide 3D data, ranging from 2 camera photogrammetry to multiple camera stereo systems.

Many earlier attempts to record 3D data on 2D sensors have been completed, and likewise many research groups around the world are currently working on camera technology but from different perspectives; computer vision, algorithm development, metrology, etc. Plenoptic or Lightfield camera technology was defined as a technique over 100 years ago but has remained dormant as a potential metrology instrument. Lightfield cameras utilize an additional Micro Lens Array (MLA) in front of the imaging sensor, to create multiple viewpoints of the same scene and allow encoding of depth information. A small number of companies have explored the potential of lightfield cameras, but in the majority, these have been aimed at domestic consumer photography, only ever recording scenes as relative scale greyscale images.

This research considers the potential for lightfield cameras to be used for world scene metrology applications, specifically to record absolute coordinate data. Specific interest has been paid to a range of low cost lightfield cameras to; understand the functional/behavioural characteristics of the optics, identify potential need for optical and/or algorithm development, define sensitivity, repeatability and accuracy characteristics and limiting thresholds of use, and allow quantified 3D absolute scale coordinate data to be extracted from the images.



The novel output of this work is; an analysis of lightfield camera system sensitivity leading to the definition of Active Zones (linear data generation – good data) and In-active Zones (non-linear data generation – poor data), development of bespoke calibration algorithms that remove radial/tangential distortion from the data captured using any MLA based camera, and, a light field camera independent algorithm that allows the delivery of 3D coordinate data in absolute units within a well-defined measurable range from a given camera.

**Keywords:** Lightfield, Plenoptic Camera, Micro Lens Arrays, Coordinate metrology, Sensitivity, Accuracy, 3D data sets.

# Table of Contents

Table of Contents .....	i
List of Figures.....	iv
List of Nomenclature .....	xi
<b>1 Introduction .....</b>	<b>4</b>
1.1. Depth recording problems using 2d photosensor .....	4
1.2. New imaging technique: Microlens array based light field .....	5
1.3. Price tag of industrial light field cameras .....	6
1.4. Low-cost light field cameras .....	7
1.5. Problem due to black-box effect .....	8
1.6. Objectives and aims .....	9
1.7. Chapter description .....	11
<b>2 Light Field Imaging .....</b>	<b>15</b>
2.1. Plenoptic function and light field .....	15
2.1.1. Theory .....	15
2.1.2. Ray-Space Representation .....	17
2.2. Plenoptic camera systems .....	19
2.2.1. Early developments .....	19
2.2.2. Methods of light field acquisition.....	21
2.3. Microlens array based light field camera .....	23
2.4. Commercially available MLA based light field cameras.....	26
2.5. Conclusion.....	27
<b>3 Light Field Acquisition Using Microlens Array Based Cameras.....</b>	<b>31</b>
3.1 Introduction .....	31
3.2 Image acquisition using conventional cameras (2D) .....	32
3.3 Image acquisition in MLA based light field cameras .....	34
3.3.1 Pinhole approximation of LF cameras.....	35
3.3.2 Light propagation through MLA .....	38
3.4 Features of MLA based light field cameras .....	41
3.4.1 Sub-aperture images .....	42
3.4.2 Post-focusing .....	44
3.4.3 All-in-focus and depth data .....	49

3.5	Potential features of MLA based light field cameras for metrology.....	51
3.6	Conclusion.....	52
<b>4</b>	<b>Parameters of Microlens Array .....</b>	<b>57</b>
4.1	Introduction .....	57
4.2	Raw image of light field camera .....	58
4.3	Geometric and light ray dependent MLA cord .....	64
4.4	Mounting issue of MLA on Photosensor .....	67
4.5	Algorithm to generate the MLA cord.....	70
4.6	Conclusion.....	78
<b>5</b>	<b>Absolute Depth Using Lytro Camera Depth Map.....</b>	<b>83</b>
5.1.	Introduction .....	83
5.2.	Features of the Lytro cameras .....	84
5.2.1.	Performance of the Lytro cameras.....	84
5.2.2.	Lytro software depth sensitivity to optical distortion.....	87
5.3.	Depth Calibration setup.....	92
5.4.	Specification of the Lytro cameras, software and external factors .....	92
5.5.	Response curve.....	96
5.6	Pixel resolution of the Lytro cameras .....	98
5.7	Result 3D Measurement .....	103
5.8	Lytro response to uniform illumination .....	109
5.9	Conclusion.....	111
<b>6</b>	<b>Absolute Depth Using Stereo View Data from Lytro Cameras .....</b>	<b>116</b>
6.1.	Introduction .....	116
6.2.	Theory of perspective views .....	121
6.3.	Order of views.....	124
6.4.	Importance of touch to focus.....	127
6.5.	Base disparity between perspective views .....	129
6.6.	Quality of complex lens systems.....	139
6.7.	Depth estimation in absolute scale.....	148
6.8.	Conclusion.....	157

<b>7</b>	<b>Calibration of Microlens Array Based Cameras</b>	<b>163</b>
7.1.	Introduction	163
7.2.	Previous work and motivation	163
7.3.	Light field calibration	165
7.3.1.	Raw light field image	168
7.3.2.	Dual views of raw image	170
7.3.3.	Preferences of calibration image	172
7.3.4.	Transformation of light field data	173
7.4.	Calculating initial calibration parameters – Step 1	178
7.4.1.	Double Pinhole Model	179
7.5.	Multi stereo lens calibration - Step 2	182
7.5.1.	Preferences of input images	183
7.5.2.	Base disparity	187
7.5.3.	Updating radial distortion coefficients	193
7.6.	Results	195
7.7.	Conclusion	200
<b>8</b>	<b>Absolute Depth from Lytro Camera Light Field Data</b>	<b>205</b>
8.1.	Introduction	205
8.2.	Selection of depth calculation method	205
8.2.1.	Ray projection method	206
8.2.2.	Disparity estimation method	207
8.2.3.	Focus variation/post-focusing	208
8.3.	Post-focusing / Refocusing	208
8.4.	Role of TTF in refocusing	213
8.5.	Depth algorithms	214
8.6.	Conclusion	220
<b>9</b>	<b>Conclusion</b>	<b>224</b>
9.1.	Introduction	224
9.2.	Key Conclusion	224
9.3.	Future work	232
<b>10</b>	<b>Appendix</b>	<b>236</b>
<b>11</b>	<b>Reference</b>	<b>246</b>

## List of Figures

Figure 1.1 Raw depth results (a) and computationally altered depth results (b) using MRF algorithm.....	9
Figure 2.1 Parameterizing a ray position (X, Y, Z), direction ( $\alpha$ , $\phi$ ), wavelength ( $\lambda$ ) and time (t)[17] .....	16
Figure 2.2 Two-plane parameterization of the light field [20] .....	16
Figure 2.3 Illustrative plots of the ray-space diagram [18].....	18
Figure 2.4 Lippmann concept of capturing Light field data (redrawn) [28].....	20
Figure 2.5 Multi-camera array system (a) and (b) used by Stanford university Multi-camera models [46], (c) Camera system used in cinematography and (d) Pelican camera [21, 22] .....	22
Figure 2.6 Code aperture model and corresponding blur pattern .....	23
Figure 2.7 Camera arrangement to capture light field data [47].....	23
Figure 2.8 Location of lenses in the plenoptic camera system [18] .....	25
Figure 2.9 Location of lenses in focused plenoptic camera 2.0 system [18].....	26
Figure 2.10 (a) Lytro first generation camera, (b) second generation Lytro Illum, (c) Raytrix camera [21, 22]. .....	27
Figure 3.1 camera model (thin lens model) .....	32
Figure 3.2 Blur-circle or Circle of confusion. ....	33
Figure 3.3 Effective boundary region classified according to the position of lenses .....	35
Figure 3.4 Light propagation through the microlens .....	36
Figure 3.5 Primary (top) and secondary (bottom) light ray models .....	37
Figure 3.6 Views of an object recorded by a single microlens.....	39
Figure 3.7 representation of distance-dependent microimage resolution and overlapping of light rays .....	40
Figure 3.8 Light field image on a 2D photosensor .....	43
Figure 3.9 2D camera views generated by Lytro Desktop software tool (top) and enlarged.....	44
Figure 3.10 Manual procedure to calculate microlens cord manually.....	46
Figure 3.11 Real and Virtual photosensor positions.....	47
Figure 3.12 Ray-space coordinates .....	48
Figure 3.13 Features of light field technology available with the Lytro family of cameras.....	50

Figure 3.14 Classification of measuring devices .....	52
Figure 4.1 Light field camera image versus conventional camera image .....	59
Figure 4.2 Different regions in a Hexagonal and Spherical MLA.....	61
Figure 4.3 FOV, angular resolution and focal length of a microlens array .....	62
Figure 4.4 Pitch of microlens array .....	63
Figure 4.5 Angle of incoming light rays.....	64
Figure 4.6 Geometric vs light directed MLA cord model .....	66
Figure 4.7 Shift map(pixels) of Geometrical cord vs light directed cord method for Illum camera .....	67
Figure 4.8 Misalignment in X and Y plane – Tilt.....	68
Figure 4.9 Perfect alignment of photosensor-MLA assembly (top) versus rotational misalignment $\theta_x$ and $\theta_y$ (bottom).....	69
Figure 4.10 Algorithm to find Automatic MLA cord using the light-directed method..	70
Figure 4.11 Representation of white image and corresponding binary image for a given threshold value .....	71
Figure 4.12 Threshold value vs Number of cord detected- Lytro-I.....	72
Figure 4.13 Threshold value vs Number of cord detected-Illum.....	73
Figure 4.14 Angle vs Cord detected – Lytro-I.....	75
Figure 4.15 Angle vs Cord detected – Illum.....	75
Figure 4.16 Representation of two cord detection results .....	77
Figure 4.17 Results of light directed cord detection for the Illum cameras white image .....	77
Figure 5.1 Depth response of Lytro camera to (a) Black-White checkerboard, (b) RGB checkerboard, (c) Lego™ bricks placed at different distances to each other and far away from the camera, and (d) Lego™ bricks placed in a compact fashion and close to the camera. ....	85
Figure 5.2 The optical distortion effect on depth map for the Lytro-I generation cameras 2D view (top) and 3D view (bottom) for a flat surface at 20 mm distance from the camera .....	88
Figure 5.3 The optical distortion effect on a depth map of the Illum camera 2D view (top) and 3D view (bottom) for a flat surface at 20 mm distance from the camera .....	89
Figure 5.4 Represents region selected for averaging technique (top).....	91

Figure 5.5 Experiment set-up to generate distance related greyscale values.....	93
Figure 5.6 Procedure to generate an average greyscale depth map .....	95
Figure 5.7 Response curves of LC1 and LC2 camera at 1400cd (A) and 1600cd (B) respectively, and LC3 camera (C).....	97
Figure 5.8 Procedure to calculate pixel resolution .....	99
Figure 5.9 Pixel resolution of Lytro cameras .....	102
Figure 5.10 procedure to generate 3D data from Lytro raw data .....	105
Figure 5.11 Raw scene data (top) and corresponding metric depth data (bottom) of the Lytro-I generation camera.....	106
Figure 5.12 Raw scene data (top) and corresponding metric depth data (bottom) of the Lytro Illum camera.....	107
Figure 5.13 Lytro-I generation (LC1) RGB scene (left) and corresponding depth map (mm) (right).....	108
Figure 5.14 Lytro-I generation (LC2) RGB scene (left) and corresponding depth map (mm) (right).....	108
Figure 5.15 Illum camera (LC3) RGB scene (left) and corresponding depth map (mm) (right) .....	108
Figure 5.16 A screenshot of depth data rendered along with the RGB value using a 3D viewer software package (MeshLab) (Lytro relative depth result and right- Z calibrated data).....	109
Figure 5.17 Results of uniform illuminations with different colour and corresponding depth maps from the Lytro family of cameras. The top row represents images of the Lytro-I generation camera and bottom row represent Illum camera.....	110
Figure 6.1 A set of three perspective images generated using the Lytro desktop software (top) compared the same perspective images generated manually (bottom) ....	119
Figure 6.2 Representation of intensity values of the scan line shown in Figure 6.1 ....	120
Figure 6.3 Disparity estimation of the Lytro images (left) and manually generated views (right). .....	120
Figure 6.4 Illustrating the internal compound lens adjustments to capture point object (subject).....	123
Figure 6.5 Results generated by Lytro Desktop software under ‘Editable Living Pictures Mode’ .....	124

Figure 6.6 Experimental setup to calculate the sequential order of the perspective views generated by the Lytro Desktop Software.....	125
Figure 6.7 Overlapped corner points of perspective views with the 00-image file as a central view for Illum camera (top row) and Lytro-I generation camera (bottom row). Numbers 00 to 06 represent the perspective views (perspective view image files) of the Lytro family of cameras considering 00 as the central view. ....	126
Figure 6.8 Representation of the sequential order of perspective images with respect to reference image marked at location (0,0) and resulting perspective orientation chart (sequential order) (bottom right).....	128
Figure 6.9 Image formation in a conventional fixed focal length camera (right) and in a microlens array based LF camera (left) (figure not to scale).....	131
Figure 6.10 Interaction of light rays in a conventional multi-camera system with the photosensor under user-selected plane at the sharp focus.....	133
Figure 6.11 Experimental setup to illustrate the importance of ‘Touch to Focus’ feature and gather the information of light propagation before and after the TTF line	134
Figure 6.12 Representation of multiple checkerboards used to select the user plane and show one of the experimental results of finding the disparity between multiple checkerboards.....	136
Figure 6.13 Disparity estimation at the focus plane in the X axis of all perspective views (Illum camera).....	137
Figure 6.14 Disparity estimation at the focus plane in the Y axis of all perspective views (Illum camera).....	138
Figure 6.15 Representation of focus plane selection by the Illum camera under user selected CB1 plane using TTF feature and resulting disparity estimation in a X direction (axis) of 01 view .....	141
Figure 6.16 Representation of focus plane selection by the Illum camera under user selecting CB3 plane using TTF feature and resulting disparity estimation in a X direction (axis) of 01 view .....	141
Figure 6.17 Representation of base plane accuracy, focus plane selection by the Illum camera under user selected CB2 plane using TTF feature, and resulting disparity estimation in a X direction (axis) of 01 view .....	142
Figure 6.18 Focus motor readings of the Lytro Illum camera focused at different distances using .....	143



Figure 6.19 Zoom motor readings of the Lytro Illum cameras for objected focused at .....	144
Figure 6.20 Focal length obtained from the Lytro Illum raw files for different values of ZS and FS.....	144
Figure 6.21 Representation of disparity values generated for the user-selected plane (CB1 plane) at different distances from the Lytro -I generation camera (disparity estimation in a X direction (axis) for 01 view).....	145
Figure 6.22 Representation of disparity values generated for the user-selected plane (CB3 plane) at different distances from the Lytro -I generation camera (disparity estimation in a X direction (axis) for 01 view).....	146
Figure 6.23 Metric depth calculation method overview represented along with disparity bar, TTF line, perspective views and shifted views .....	147
Figure 6.24 (a) Disparity estimation of a scene captured using Lytro Illum, (b) results for images without row-shifting and (c) the result obtained for images with row shifting .....	148
Figure 6.25 Algorithm for calculating absolute depth data using base disparity estimation.....	150
Figure 6.26 Disparity generated between 00 and 01 views .....	151
Figure 6.27 Absolute depth generated using method-1 for the Illum cameras.....	153
Figure 6.28 Absolute depth generated using method-2 for the Lytro-I generation cameras.....	154
Figure 6.29 work volume illustration using checkerboard plane for TTF at (a) CB1, (b) CB2 and (c) CB3.....	155
Figure 7.1 LF raw image generated using toolbox (left) versus sub-aperture view generated using LDS (right) of an object at 200 mm from the camera. ....	169
Figure 7.2 Representing telescopic and binocular views by the Lytro family of cameras (Illum) at 200 mm (top row, left) and 800 mm (top row, right) focal plane from the camera .....	171
Figure 7.3 Representing telescopic and binocular views by the Lytro I-generation cameras at 200 mm (top) focal plane from the camera and corresponding enlarged views (bottom).....	172
Figure 7.4 Representation of binocular and telescopic view exhibited by Lytro family of cameras defined using a checkerboard image.....	173

Figure 7.5 Light field data and sub-aperture camera views.....	176
Figure 7.6 Double pinhole model .....	179
Figure 7.7 Near and far regions of raw LF image representing the distribution of light rays when .....	184
Figure 7.8 Representing control points in a sub-aperture and corresponding synthetic data to generate target plane parallel to the camera plane .....	185
Figure 7.9 Representing input image base accuracy score for the Lytro family of cameras.....	185
Figure 7.10 Represents a group of Neighbourhood (top), PR,SR and TR for Illum (a) and.....	186
Figure 7.11 Base disparity estimation sketch .....	188
Figure 7.12 Algorithm for base disparity estimation.....	188
Figure 7.13 Graphs showing the distribution of the feature points in the various neighbourhood .....	191
Figure 7.14 Graphs showing the distribution of the feature points in the various neighbourhood .....	192
Figure 7.15 Base disparity estimation of different neighbourhood groups for the Illum camera .....	193
Figure 7.16 Base disparity estimation of different neighbourhood groups for the Lytro-I camera .....	193
Figure 7.17 Feature points of the calibrated image (step-2).....	194
Figure 7.18 Mean re-projection error of sub-apertures using 35.1 mm grid generated using previous methods.....	197
Figure 7.19 Mean re-projection error of sub-apertures using 35.1 mm grid-Lytr-I ...	197
Figure 7.20 Mean re-projection error of sub-apertures using 35.1 mm grid-Illum .....	198
Figure 7.21 Disparity estimation of Lytro data before calibration, illustrated using 3D mesh using MeshLab software.....	199
Figure 7.22 Disparity estimation of Lytro data after calibration, illustrated using 3D mesh using MeshLab software.....	199
Figure 7.23 LF data after removing distortions .....	200
Figure 8.1 Refocusing technique and ray tracing using two lenslets.....	206
Figure 8.2 Necessary changes in a conventional imaging system to record different ..	209
Figure 8.3 Refocusing at different plane/distances from the camera .....	211

Figure 8.4 Refocusing algorithm .....	212
Figure 8.5 Focus/sharpness algorithm .....	213
Figure 8.6 Image considered for alpha verses TTF experiment .....	213
Figure 8.7 Alpha value compared with the TTF of the Lytro cameras .....	214
Figure 8.8 Procedure to generate absolute depth data using disparity method.....	215
Figure 8.9 Algorithm for depth calculation using disparity method.....	216
Figure 8.10 Normalized disparity estimation(right) and image considered (left) .....	218
Figure 8.11 Relative plane depth map (mm) .....	218
Figure 8.12 Top view of the result (mm).....	219
Figure 8.13 Absolute depth map relative to user-defined focus plane distance .....	219
Figure 8.14 Focus variation results for different values of alpha.....	220

## Nomenclature

LF	: Light Field
NA	: Numerical aperture
MLA	: Microlens array
LC1, LC2	: Lytro-I generation cameras
LC3	: Lytro Illum camera
TTF	: Touch-to-focus
CLS	: Compound lens system
$c_x, c_y$	: The optical centre of the image
$r$	: The distance of image pixel $(x, y)$ from reference point
$d_{Avg}$	: Average depth map
AZ	: Active Zone
IAZ	: Inactive Zones
ROI	: Region of interest
$K1, K2, K3$	: Radial parameters of calibration
$P1, P2$	: Tangential parameters
$c_h, c_w$	: Height and Width of the checkerboard
$n_h, n_b$	: Distance of the checkerboard from the camera
$h$	: Distance between optical axis of microlens to the edge of microimage
$d''$	: Distance between microlens plane and the photosensor
$\bar{f}$	: Focal length of the microlens unit/ entire MLA
$p_v$	: Number of perspective views
$P_x$	: Microlens pitch in $x$ direction

$P_y$	: Microlens pitch in $y$ direction
$[X, Y, Z]$	: World coordinate system points
$[s, t]$	: Main lens plane
$[u, v]$	: Microlens plane
$[x, y]$	: Photosensor plane
$[m, n]$	: Pixel count in the photosensor
$LF$	: Raw lightfield image
$lf$	: Transformed lightfield image to our standards
$\varphi_i$	: Individual camera view
$L_{mm}$	: Microlens addressing mode
$b_p$	: Base plane accuracy
$ZS$	: Zoom motor step values
$FL$	: Focus motor step values
$b_{dx}$	: Base disparity values
$d_{map}$	: Disparity map
$d_{mm}$	: Absolute distance calculated from the disparity map
$Z_{mm}$	: Focal plane distance
$L_{dm}$	: Direction addressing mode
MLA-Ps	: MLA-Photosensor gap (mm)



# 1



## INTRODUCTION

## *Overview*

The research reported here in this thesis considers low-cost Light field (Plenoptic) camera technology and the potential relevance to coordinate metrology for industrial applications. The research investigates commercial domestic grade light field cameras, develops algorithms and processes for data manipulation and calibration of commercial camera systems, resulting in absolute depth maps in SI units.

In this thesis, the light field camera and related technology is explained in a progressive fashion. The early chapters deal with history, alternative techniques for coordinate measurement and development cycle of light field cameras. Later in Chapter 4, basic optical elements used in light field cameras are explained, along with necessary calculations to find the basic specifications and parameters of the optical elements. The initial results suggesting the potential use of light field cameras in coordinate metrology is described in Chapter 5, where greyscale depth maps from the Lytro family of cameras is used to provide depth information, using a single light field image. The advantages/disadvantage and features of the Lytro family of cameras are illustrated in Chapter 6, where Lytro-I generation cameras are compared with the Lytro Illum. To overcome the black-box effect, an algorithm is developed, and absolute depth is calculated using calibrated light field images. The calibration algorithm is discussed in Chapter 7 and implementation is defined and illustrated in Chapter 8.

The main contributions of this thesis are highlighted in Chapter 9, where the novelty of every Chapter is rated between 0 to 5. The conclusion of this research is also discussed in Chapter 9 along with recommended plans and ideas that can be conducted to further develop the research in the field of metrology using light field cameras especially using low-cost Lytro cameras.



# 1 Introduction

## 1.1. Depth recording problems using 2D photosensor

The introduction of digital cameras, during the 1960-70's, was a major milestone in the field of photography. Since that point, digital cameras have been and are used in almost every field ranging from astronomy to taking a selfie. These digital cameras are highly efficient in recording light information, passed on to the photosensor using suitable optical lens systems, and store the data in a digital format (images). Hence, these cameras have made their way into many engineering fields, where observation, recording and storing activity was the key.

A new field of engineering emerged based on digital images and videos captured using digital cameras, known as computer and industrial vision. A rapid growth in the number of applications (especially related to engineering) depended on digital cameras, exposed some of the downsides of this technology. Digital cameras use a single 2D photosensor to record the light rays, thereby efficiently recording height ( $x$ ) and width ( $y$ ) of an object but fail to register the 3<sup>rd</sup> dimension (depth,  $z$ ), and these cameras are considered as conventional cameras in this thesis. However, the captured images represent the third dimension (depth) by changing the object size, i.e. objects close to camera appears bigger compared to the object at the far distance.

Many efforts were made to record the 3<sup>rd</sup> dimension on the 2D photosensor, with one alternative approach being to add additional pixel bits under the existing pixels to capture depth information [1]. But this technology required developing a new photosensor, that would significantly escalate the price. Some of the successful techniques related to conventional cameras were:

1. Capturing a sequence of images by varying the distance of the camera from the object of interest provides depth information [2], [3], by comparing the size of the object and other camera parameters (focal length, pixel count). The changes made with camera distance must be measured accurately to generate good depth results in absolute units; otherwise, the depth information would be relative to the scene.

2. Using two/multiple 2D photosensors to capture the parallax information of the scene [4]. This undoubtedly (since human vision works with a pair of eyes) provides acceptable depth information but involves, a few drawbacks related to; positional relationship of the individual camera with respect to other camera/s, and necessarily repeating this relationship step for every new setup.

These methods led to other techniques such as calculating focal error, focal gradients [5] [6], intensity variation, perspective and stereo-vision [7]. More advanced techniques used digital cameras with a laser [8], [9], and fringe projection [10], [11] to record the depth.

However, the solutions of capturing depth information using the digital photosensor still consist of multiple devices and calibration of these devices to perform as a single unit has been difficult. Many experiments have been carried out to date to make/improve depth registration as simple as capturing a photo, using a single photosensor.

## **1.2. New imaging technique: Microlens array based light field**

Light field (also known as Plenoptic imaging) has emerged as an interesting technique, with promising features to record depth information indirectly on a single photosensor. In simple terms, the light field (LF) can be defined as a method of capturing full 4D radiance information of light rays (direction of light rays from 2 lens elements). The concept of LF imaging originates as early as 1908, yet it has been in the past few decades that this technology has been considered for practical application.

LF cameras are very similar to conventional cameras in design but consist of additional optical elements, which enable these cameras to record light information in a different fashion when compared with conventional cameras. The recorded image can be decoded to extract the direction of light rays that travelled into the camera space and registered on a single photosensor to provide the intensity information. There are several methods of capturing LF data on a 2D photosensor, but one of the core ideas to this thesis is to restrict the emphasis of the research to microlens array based LF acquisition techniques. With this scope, it is more straightforward to pinpoint the technology referred to in this thesis, amongst the field of LF imaging techniques.

Capturing light data on a single photosensor to generate the depth information would overcome many drawbacks of the techniques briefly identified in Section 1.1. The microlens array based LF technique opens the potential for new solutions to many applications, where conventional cameras fail to provide necessary information with a single photosensor.

Computer vision groups are actively exploring this new branch of imaging and some noticeable results have been published already, proving and validating the LF concept. Whilst in other engineering fields, especially in metrology, limited work related to LF imaging has been published. In addition, for the majority of published work, the LF generated depth maps are in relative scale (data represented in relative grey scale values) rather than absolute scale and no information related to important key terms (depth accuracy, repeatability) that defines an instrument were disclosed.

Hence, the primary goal of this work is to understand the principles and the concepts of LF imaging from an engineering point of view with reference to potential metrology use. Many existing research papers and patent applications reveal the technology in a top-down fashion, i.e. the LF imaging is described without enough information to understand the function of every optical element in the camera system. Hence in this work, a bottom-up approach is considered, i.e. the concepts are illustrated using ray diagrams, and simple point source objects to give a better understanding of what really happens inside the LF cameras. Also, this work aims towards improving the general understanding of all components necessary for the camera to exhibit LF imaging. This work also considers the use of commercially available LF camera models (specifically consumer grade cameras), since few companies have managed to design and develop LF cameras targeting distinctive groups.

### **1.3. Price tag of industrial light field cameras**

Lytro and Raytrix are two companies, who have launched microlens array based LF cameras that are compact, just like DSLR cameras. Lytro cameras are targeted for the consumer market, while the Raytrix cameras are restricted to engineering applications. Both these cameras are designed and integrated with all required optical elements within the camera body, enabling ease of use. It is evident from the patent applications [12], [13] that both these cameras work under the same LF principles, yet a high-cost difference can be noticed. The price difference is in the ratio of 1:70+ for the Lytro

cameras and the Raytrix cameras respectively. With such high prices, it can be difficult to procure expensive cameras for all applications. The accuracy and other results expected from the cameras differ with applications, i.e. object detection on conveyer belt application would require depth accuracy in the order of millimetres, while micrometre accuracy for other applications. Hence using compact, single unit, camera technology for all applications will not suit because of huge investment, while the LF technology supports both these applications.

A large investment for LF cameras specifically designed for engineering applications potentially limits the widespread acceptability and availability of this interesting technology. So, this issue drives the second aim of this thesis, to bring down the initial investment required for using LF cameras to generate depth information. This work investigates low-cost LF camera models, to see if they have potential to be used in engineering application to provide absolute depth results.

#### **1.4. Low-cost light field cameras**

In contrast with the expensive LF camera available out in the market, there are low budget cameras that follow the basic principles of the LF camera model. The Lytro company introduced two versions of LF cameras, the Lytro-I generation and the second generation Illum, that are approximately £100 and £1,000 respectively. Despite being cheap (in comparison with Raytrix), these cameras record a promising amount of radiance and record additional information of incoming light as seen with expensive cameras.

The Lytro cameras being cheap, come with drawbacks corresponding to the way the results are generated and exhibited. With respect to metrological applications, the Lytro cameras provide depth maps in relative greyscale to the scene, i.e. the depth map is defined in terms of varying greyscale values (0 to 255) rather than absolute units (e.g. millimetres) as expected by many engineering applications.

The computer vision community has generated many new methods of improving the depth results generated by the Lytro cameras, but very limited results have been published where early steps towards converting relative scale depth maps into absolute scale values can be noticed. For this reason, the Lytro cameras have not been

considered by engineering community (where depth maps need to be in absolute scale) for practical applications.

Potentially this has limited the use of LF imaging method in the industrial environment due to high price (Raytrix cameras) and lack of desired results (Lytro depth maps). This issue has been considered in this thesis as major work to link these limitations, by developing a broad understanding of the Lytro cameras and providing a novel model that generates absolute depth information using low-cost LF cameras.

### **1.5. Problem due to black-box effect**

Some new instruments/devices used in engineering applications provide efficient results but hide the details of process and techniques used to provide results. If the users are provided with ‘input-process-output’ cycle, it will be easier to understand the nature of the device and judge the quality of the results produced. The lack of ‘input-process-output’ cycle is considered as a Black-Box Effect in this work and this effect is very applicable to the Lytro cameras.

The Lytro cameras work on the LF but the actual design of how each component in the camera supports capturing light radiance is not disclosed (not publicly available). Some of the major key observations pointing the Lytro cameras exhibiting black-box effect are:

1. The Lytro company encode the images captured using their cameras in .LFP/.LFR formats, making the files not usable directly. Hence to use the raw sensor data one must find a way to decode these files.
2. Lytro cameras are supplied with a software package that consists of different ways of accessing results from captured data, for example, raw metadata file, all-in-focus, depth map. So, to know if any of these results are useful and help in reaching our goal, it is necessary to complete experiments and understand what the results mean with respect to the captured data.

Hence, to use low-cost Lytro devices for engineering application, first, the nature of camera must be identified, i.e. details of hardware, software, working environments, build quality and much more, depending upon the application requirement. Thus, part of this thesis details the development and understanding of the concept of LF imaging

and relating to the Lytro family of cameras. By conducting experiments, the nature of the Lytro family of cameras is understood before developing into practical applications.

The Lytro cameras are commercially available and designed for a general audience rather than engineers, not especially targeting metrological applications. Hence in this work, the raw and unprocessed results are presented to better understand and find the relationship between the camera and corresponding depth results. Computer vision groups have previously provided results from Lytro cameras that are computationally modified to fit the scene data, i.e. by applying graph theory algorithms, for example by using the Markov Random Field (MRF) [14], [15], where discontinuous regions are modified into the smooth and continuous region, as shown in Figure 1.1. In engineering applications, especially metrology, a pixel in the original depth map represents true behaviour of the system and depending upon the application requirements computational algorithms are used to generate smooth variations between neighbouring pixels. As discussed earlier, computer vision groups have been publishing results similar to Figure 1.1 (b), while the true results are as shown in Figure 1.1 (a).

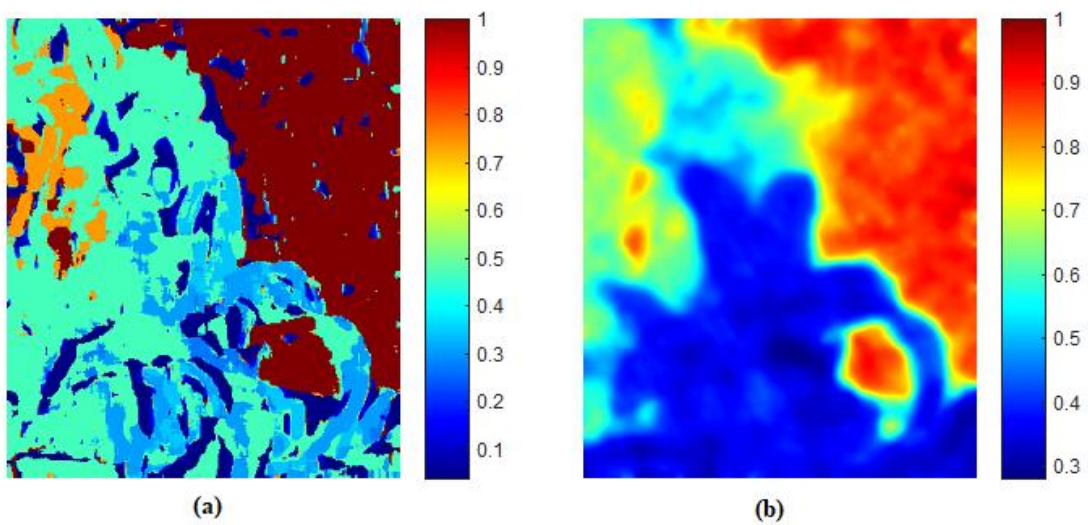


Figure 1.1 Raw depth results (a) and computationally altered depth results (b) using MRF algorithm

## 1.6. Objectives and aims

The objectives and aims of the research reported here in this thesis can be summarised as:

- To understand the concept of LF imaging from a metrological point of view.

The understanding is not just limited to the theory and principles of LF imaging

but also includes acquiring knowledge of hardware components, optical elements, especially the micro lens array (MLA), and the software.

- Use a commercially available low-cost, microlens array based LF camera (Lytro) to familiarise and get hands-on experiences with LF cameras and eventually identify the suitability of the LF cameras in engineering applications.
- To understand the hardware and software package provided by the Lytro company/available elsewhere, that promotes usage of the Lytro cameras in engineering applications.
- Identify (experimental) any distinctive features of the Lytro cameras with respect to other LF cameras available commercially. Thus, proving an initial advantage for the Lytro cameras for engineering applications.
- Identify potential key problems related to the Lytro cameras preventing them to be used in engineering applications. Also, provide novel work models (if any) to overcome these potential disadvantages and demonstrate in practical applications.
- Provide novel, a systematic review of quality of the Lytro family of cameras with reference to engineering applications.
- Develop a novel model that uses the Lytro camera just as an LF input source and the rest of the processes are transparent to the user to avoid the Black-Box Effect. The algorithms in the model should be compatible with other single focal length MLA based cameras or multi-camera array cameras capable of capturing LF data.
- Develop novel camera calibration techniques that can be used to calibrate the raw images of the Lytro cameras and other single focal length MLA based LF

cameras by improving the time required for each calibration cycle and overall calibration score (RMS re-projection error, see 7.3).

- Develop a novel algorithm to provide depth data in SI units using the features of the Lytro cameras.

## **1.7. Chapter description**

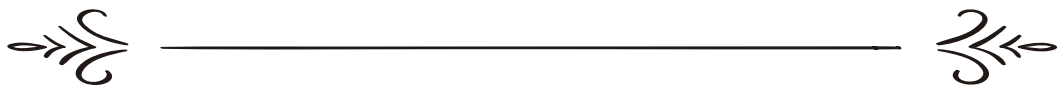
The structure of this thesis can be briefly summarised in the following bullet points:

- Chapter 2 introduces the history, LF techniques and different approaches to capture light radiance.
- Chapter 3 specifically targets the LF cameras based on the microlens arrays (MLA) and illustrates different features offered by LF techniques for metrological applications.
- Chapter 4 illustrates the interaction of light rays with the MLA and important key measurements necessary to access LF features.
- Chapter 5 identifies the Black Box Effect in detail corresponding to the Lytro camera and proposes novel absolute depth generation model using greyscale data from the Lytro cameras
- Chapter 6 introduces sub-aperture images generated by the Lytro cameras and the novel concept of generating absolute depth using these images. Also, this chapter identifies the build quality the Lytro-I generation and Illum cameras.
- Chapter 7 describes calibration of the raw LF images of the Lytro camera, to remove lens distortions using sub-aperture images by considering novel virtual lens techniques and quantifies the calibration process using key parameters obtained from the MLA-photosensor assembly.
- Chapter 8 develops a new depth algorithm to generate absolute depth data using disparity techniques and explains the major drawbacks of the Lytro family of cameras preventing usability in engineering applications.
- Chapter 9 summarises the work completed, identifies the novel contributions of this work and points towards future directions of research.





# 2



## REVIEW OF LIGHT FIELD TECHNOLOGY

## *Overview*

The method of capturing light rays using photosensitive pixels for recording data has been used in practical applications for a long time in engineering applications. But capturing the depth data on the same photosensor has been challenging and many efforts have been made.

Many new techniques have been proposed to capture the depth data using digital cameras. Few of them use some additional optical elements or resize the existing components to achieve the goal of capturing 3D information.

This chapter introduces to the field of light field (LF) imaging by narrating the history, the inventor of LF technique and early methods used to capture additional data. Also, illustrate LF imaging from a broad perspective and later pinpoints the method used in this thesis. Some of the alternative techniques to capture LF data are highlighted toward the end of the chapter.

This chapter plays an important role to make understand the basic terminology and mathematical functions used to address LF data and link chapter 3, where LF imaging techniques will be narrowed on to MLA based LF imaging.

## 2 Light Field Imaging

Imaging devices play a vital role in the modern engineering applications by recording light intensity using light-sensitive materials. The method of registering photons has advanced from photo-sensitive plates to compact size digital photosensors. The advancements are not just limited to the way of recording the photons but have gone further in making use of these photons. Until recently, imaging devices were considered as a static recording device, since there was no way to alter the information recorded by the imaging devices once the photons were recorded. Thus, a blur region in the image stays blur without any option to bring the region back to sharp focus. Many optical lens arrangements have been suggested to capture additional information along with registering photons on the photosensor pixels. Light field imaging is one of these methods, which allows the user to record directions of the light rays using optical lenses and hence virtually alter the light rays after recording on the photosensor.

### 2.1. Plenoptic function and light field

#### 2.1.1. Theory

Light rays [16] carry a vast amount of measurable information that is captured as light intensity using imaging devices. The main goal of these imaging devices is to record details of light rays in a way that this information can be recalled when necessary (digital version as an image). But many modern imaging devices are not fully equipped to preserve complete information (depth, polarity) carried by light rays into imaging devices thereby lacking the ability to produce complete information of the scene which the user intended to capture.

The light rays are the primary means of transferring various information of the object/scene onto an imaging device. These light rays can be used to define the 3D world by considering a Plenoptic function. The Plenoptic function is a 7D function that models a 3D dynamic environment by recording the light rays at every space location  $(X, Y, Z)$ , towards every possible direction  $(\alpha, \phi)$ , over any range of wavelength  $(\lambda)$  and at any given time  $(t)$  [17], [18] given by Equation 2.1.

$$LF = l(X, Y, Z, \alpha, \phi, \lambda, t) \quad 2.1$$

The full complexity of the optical information filling space can be formalized as a single function (Equation 2.1), which Adelson and Bergen [19] called the “*plenoptic*

function”, derived from the Latin root *Plenus*, which means complete or full, and *optics*, which means pertains to vision, which records the appearance of the whole world as shown in Figure 2.1 [17].

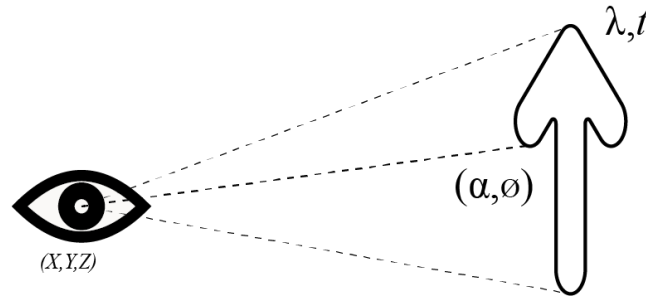


Figure 2.1 Parameterizing a ray position  $(X, Y, Z)$ , direction  $(\alpha, \phi)$ , wavelength  $(\lambda)$  and time  $(t)$ [17]

The dimensionality of the plenoptic function can be reduced by considering a single frame of the image, i.e. a single value for  $t$  in Equation 2.1, with monochromatic light rays (constant  $\lambda$ ). The radiance along the light ray is constant, only two parameters are needed for the location, effectively omitting  $Z$ . Furthermore, it is common to replace the angular coordinate values  $(\alpha, \phi)$  with cartesian coordinates  $(u, v)$  and the viewing position  $(X, Y)$  can be replaced with recording location of the imaging device  $(x, y)$ . As shown in Figure 2.2, the light rays can be assumed to pass two plane systems in the path from left to right. The light ray intersects the first plane at coordinates  $(u, v)$  which govern the location of the ray; it continues to travel and intersects the second plane at coordinates  $(x, y)$  effectively specifying the propagation direction.

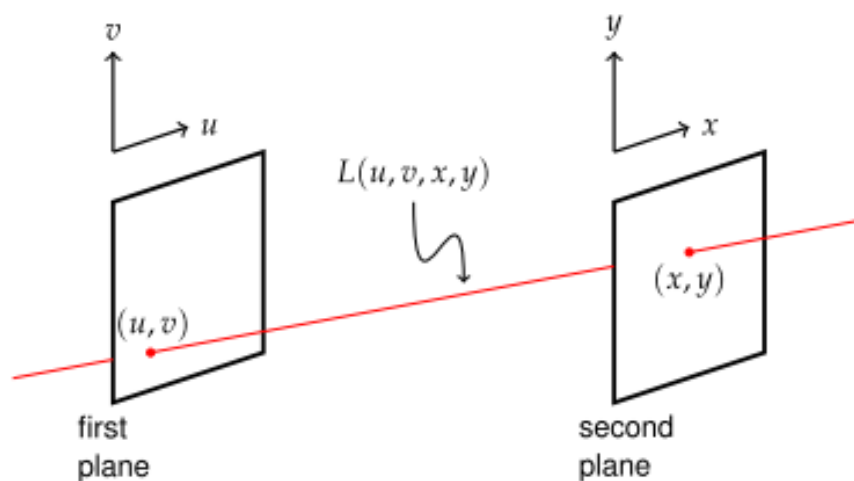


Figure 2.2 Two-plane parameterization of the light field [20]

Instead of calling Equation 2.1, the plenoptic function, it is commonly called the “lumigraph” in the computer graphics community [20] to refer to this 4D simplification as the LF, denoted by Equation 2.2 representing the amount of light, known as the radiance of the monochromatic light ray. When considering colour image formation, it can be assumed that the LF is a three-component vector with radiance information on the red, green, and blue components or any corresponding wavelengths [18].

$$LF = L(u, v, x, y) \tag{2.2}$$

### 2.1.2. Ray-Space Representation

Visual representation of light rays in the LF can be achieved by plotting the ray-space diagram [20, 21], which allows insight about the different collection of light rays in a compact manner. The two planes shown in Figure 2.2 can be represented as 2D diagrams (Figure 2.3), where the LF is shown by the plot of  $x$ , the spatial axis, versus  $u$ , the directional axis. The light ray  $L(u, x)$  is then represented by a point in the ray-space diagram. Figure 2.3a (left) represents the set of spaced radiance from the  $u$  plane onto the  $x$  plane, generating a regular ray-space matrix (Figure 2.3a (right)). When light rays are arranged to focus on a single point on the  $x$  plane, the ray-space diagram shows a vertical line (Figure 2.3b), indicating the light rays of different directional values are registering on the same location. In Figure 2.3c, the set of rays at different points on the  $u$  plane converge to a point at the  $x$  plane, but when the  $x'$  plane is introduced close to the  $u$  plane, the ray-space shifts from a vertical line to a tilted line (to right side) indicating different spatial locations. Another way to describe the phenomenon is that the ray-space has undergone a shearing, this being a fundamental concept to a LF camera. In Figure 2.3d, a collection of rays at different points on the  $u$  plane converge to a point at the  $x$  plane, but when the  $x''$  plane is introduced at a far distance after the  $x$  plane, these rays are again diverging to different spatial locations. Hence in the ray-space diagram, they form a tilted line (towards the left side) indicating shearing of the ray-space but to a different direction.

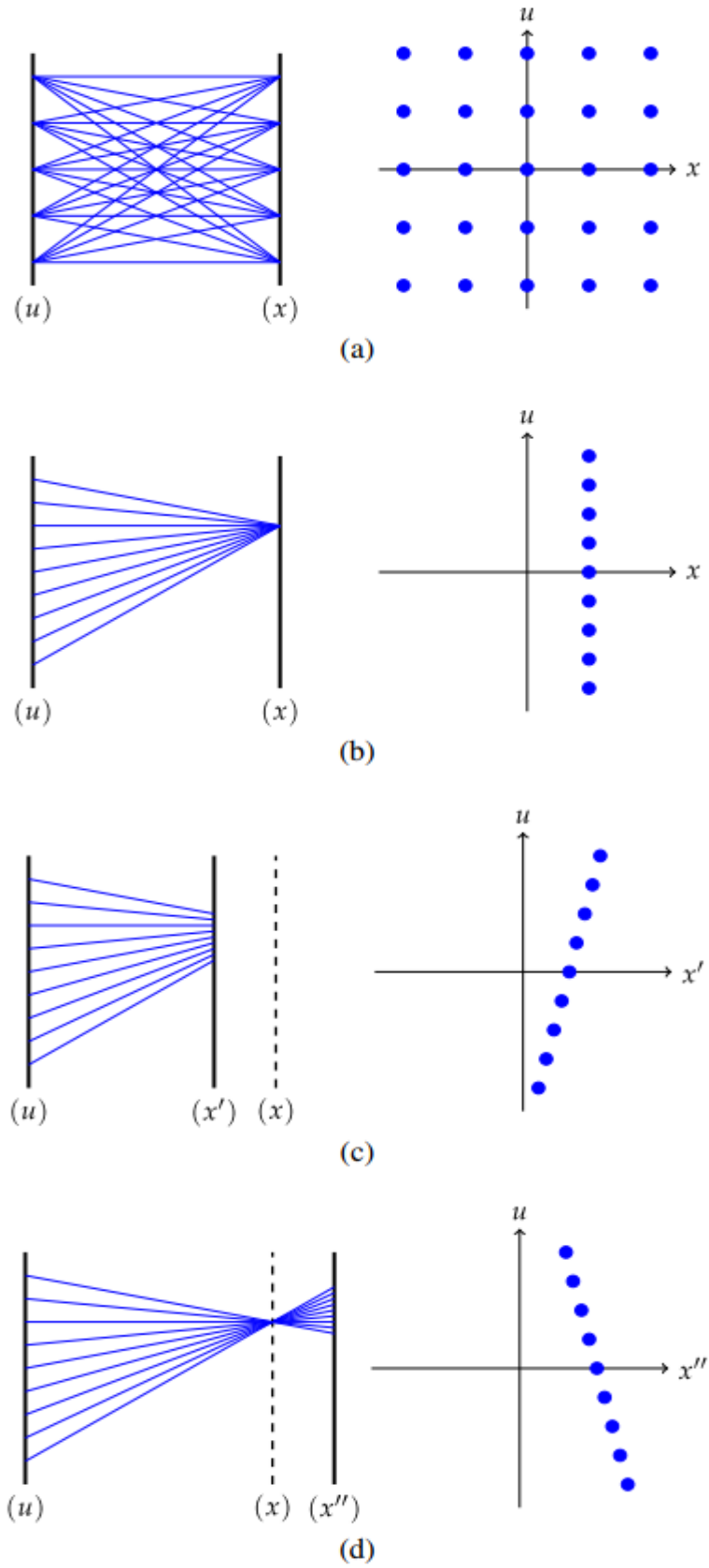


Figure 2.3 Illustrative plots of the ray-space diagram [18]

This indicates that, if the  $(x, y)$  plane is replaced by a photosensor, it would capture the irradiance of the light from the  $(u, v)$  plane. This is a summation of the incident radiance [22, 23] representing integration of the LF, mathematically given by Equation 2.3, where  $E(x, y)$  can be considered as the final image intensity on the photosensor and  $Z$  is the separation between  $(u, v)$  and  $(x, y)$ . Equation 2.3 thus provides the linkage between the 4D LF and the 2D image capture in a conventional camera: the image is a projection of the LF onto the  $(x, y)$  axes [18].

$$E(x, y) = \frac{1}{Z^2} \iint_{-\infty}^{+\infty} L(u, v, x, y) dudv \quad 2.3$$

## 2.2. Plenoptic camera systems

### 2.2.1. Early developments

The first conceptual model of recording parallax information (parallax is a displacement or difference in the apparent position of an object viewed along two different lines of sight, and is measured by the angle or semi-angle of inclination between those two lines) [25, 26] on a photosensitive plate was proposed by F.E.Ives in 1903, where a lens of diameter of 3 inches was used with aperture holes representing the pinhole equivalent mechanism. The light rays emerging from this pinhole assembly were later captured by a photosensitive plate (9). The concept of capturing light rays of objects with multiple apertures and multiple tiny lenses are illustrated in patent [27] which serves as the start of new technology.

The idea of gathering depth data using a single 2D sensor was proposed by Gabriel Lippmann in 1908 [28], entitled *Integral Photography*. It was a method of capturing LF data with the help of very small lenses similar to the primary lens. According to his idea, light rays from a point source can be recorded as multiple images and each image represents light travelled from the same light source but each light ray is shifted slightly with each other (Figure 2.4)[28]. Lippmann's experimental results included crude integral photographs made by using a plastic sheet embossed with a regular array of the microlenses, or by partially embedding very small glass beads, closely packed in a random pattern, into the surface of the photographic emulsion.



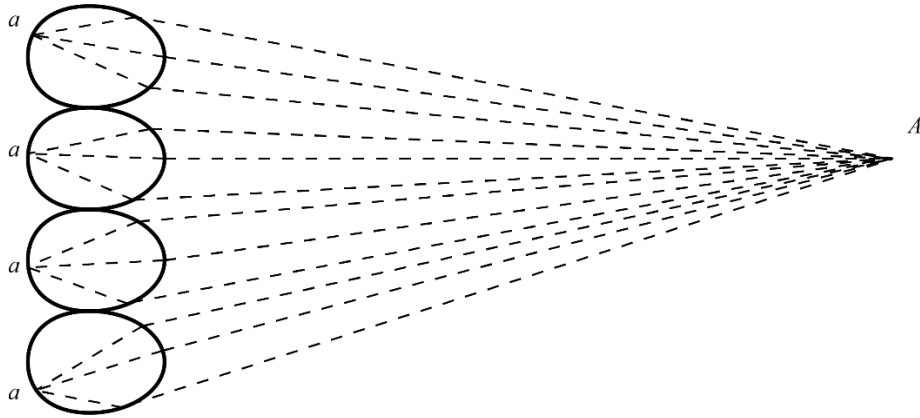


Figure 2.4 Lippmann concept of capturing Light field data (redrawn) [28]

Unfortunately, Lippmann's idea was not practically implemented for many years, primarily because of the technical difficulties in creating the lens array. An alternative technique similar to Lippmann was parallax stereograms developed by Ives in 1930 [26, 27] by adding two aspects. First, he added the main lens with a large diameter [31]. Second, he analysed the original approach and found it to be pseudoscopic (a binocular optical instrument/phenomenon that reverses depth perception) rather than stereoscopic (a method that presents two offset images separately to the left and right eye of the viewer). He then proposed a secondary exposure to invert the depth [32].

The next significant phase of development emerged primarily from the computer vision and computer graphics communities. Soon after defining the plenoptic function [25, 30], Adelson and Wang outlined an image capture system to collect such data and named it the plenoptic camera [34]. The plenoptic function describes the LF as a set of rays intersecting two 2-D planes enabling image acquisition, reconstruction of spatial and angular light information. In 2000, Isaksen [35] explored the capability of refocusing using viewpoint images captured by an array of cameras.

Advancements in precision optics and manufacturing techniques led the way to the fabrication of very small lenses with high accuracy and precision eventually leading to the development of the plenoptic imaging devices. Ng Ren carried this technological idea further to investigate refocusing based on a hand-held Standard Plenoptic Camera (SPC) having an MLA attached to the front of the sensor [21]. Afterwards, this plenoptic setup was implemented in a microscope by Levoy [36] and has been recently advanced by Broxton [37]. In 2009, Lumsdaine and Georgiev [38] improved the effective spatial resolution by introducing a new rendering technique for the Focused

Plenoptic Camera (FPC) allowing for different positions of the MLA. Nevertheless, the FPC inherently causes a loss of angular information resulting in a trade-off between angular resolution (variation in pixel information available in the sequence of images of the same object captured at a different angle) and spatial resolution (total pixel count available in the 2D image) [39]. An early investigation of depth measuring using a disparity analysis of integral images was conducted by Wu and Chun.Hong [40]. The challenge with this approach is to minimize the error resulting from the disparity map (see Sec 6.1)[16] which is due to the relatively small baseline compared to a camera array setup. Improvements have been made by Bishop [41] and Perwass [42]. The research examining the position of virtual microlenses was undertaken by Georgiev [43]. Therein, principal plane calculations of the FPC identified that virtual lenses are projected into object space comparable to an array of cameras.

### **2.2.2. Methods of light field acquisition**

Before miniature optics like the microlens was manufactured, some research groups used conventional cameras to capture LF information of the scene. Here in this system, each camera records light emitted from the scene at different angles. The amount of angular information recorded by this system is higher due to the large space between camera array systems (Figure 2.5 a and b). Theoretically, the amount of data, spatial resolution and angular resolution are higher in the conventional method of capturing LF data, whereas the huge size, complex camera control and non-flexibility are some of the down-sides of this system. The major disadvantage of the multi-camera model is the amount of time required to set up the cameras and interconnecting them under a single control unit. These days the multi-camera array model is widely used in film and cinematography fields to generate a 3D effect, still pictures and other visual effects. In this camera system, the object is surrounded completely ( $360^\circ$ ) as shown in Figure 2.5(c) to capture data from all possible angles. Since the conventional cameras can be manufactured in a compact size, the Pelican company exploited this advantage and designed an array of cameras to fit a mobile handset as shown in Figure 2.5(d), thereby providing customers with additional features than just 2D images [21, 22].

The Coded Aperture method utilizes the radiance captured by the conventional camera, by creating pre-determined, different shape and size hole at the aperture to generate pinhole effect on the photosensor. The circle-of-confusion (see Sec 3.2)[16] commonly

seen in the conventional cameras is utilised by this method to generate depth information for a short range. But a particular coded aperture mask (see Figure 2.6) works best for a specific distance scene [44]. Figure 2.6 represents a conventional camera with a coded mask and corresponding blur pattern generated when exposed to light. Similar methods to capture radiance by generating an array of aperture holes and rotating with the camera optical axis as a pivot to capture angular information has results comparable with coded aperture method. But due to the addition of the rotating mechanism within the system, this technique has not been used commercially [45]. These methods provide an efficient way of capturing light radiance, but the hardware and apparatus required to generate additional information make these systems less advantageous compared to microlens array-based devices.

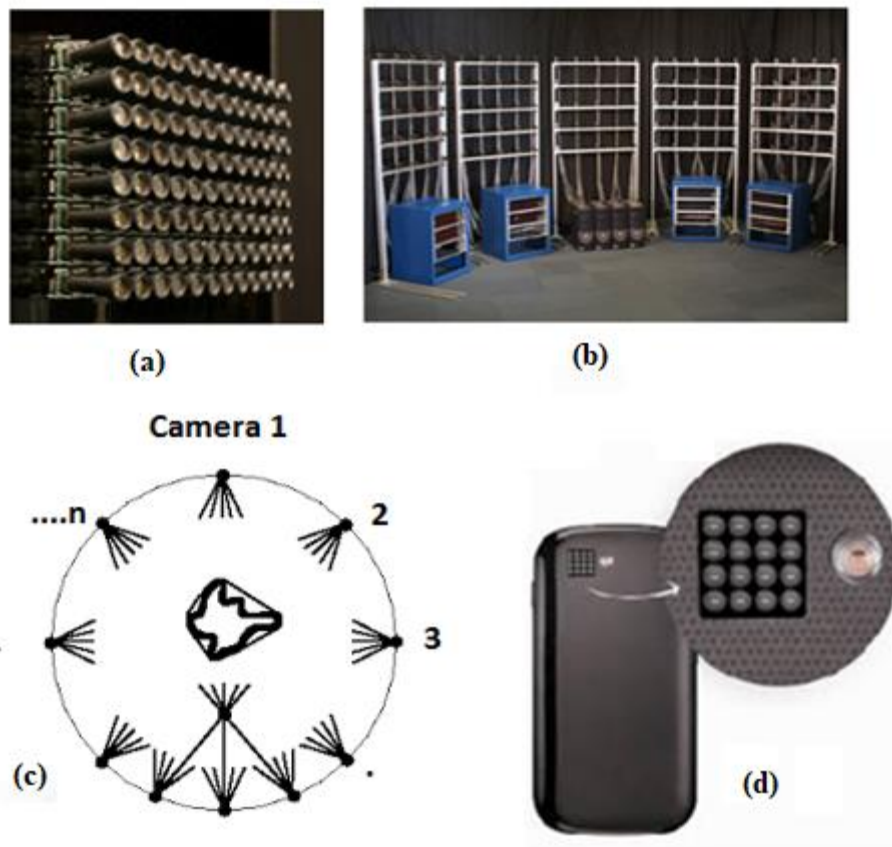


Figure 2.5 Multi-camera array system (a) and (b) used by Stanford university Multi-camera models [46], (c) Camera system used in cinematography and (d) Pelican camera [21, 22]

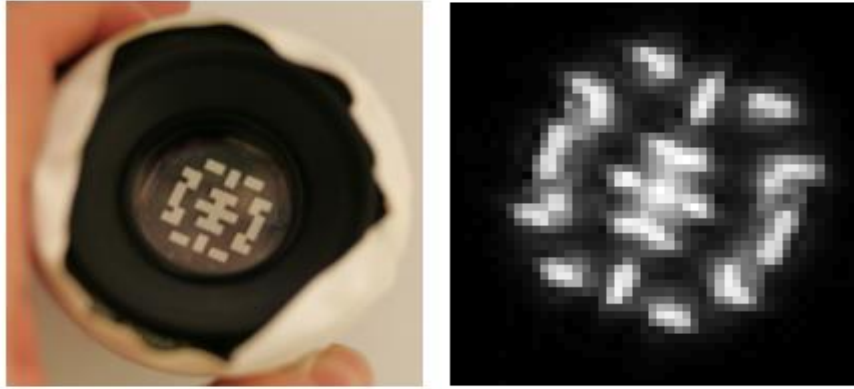


Figure 2.6 Code aperture model and corresponding blur pattern

### 2.3. Microlens array based light field camera

The advancement of precision lens fabrication techniques led the way for compact design to capture light radiance. Adelson and Wang proposed the model which was based on their single lens stereo principle and would acquire LF image information in a single image capture. In an ordinary camera, all the light striking a given photosensor element (one cell of a CCD array) is treated in the same way. The photon responses are summed, regardless of the angle of incidence. If the amount of light striking the photosensor from different directions is recorded, then it can be determined how much light came from the various sub-regions of the lens aperture.

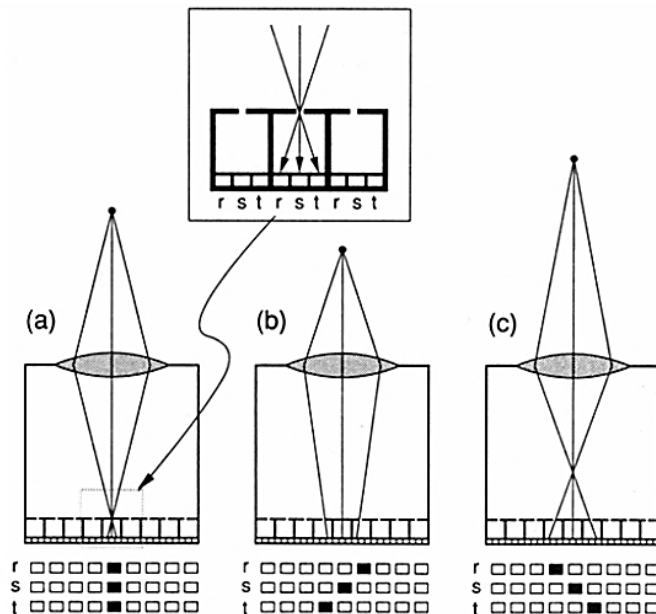


Figure 2.7 Camera arrangement to capture light field data [47]

Consider the arrangement shown in Figure 2.7, where the photosensor is covered with an array of tiny pinhole cameras. The light rays entering each pinhole camera is broken

up into three subparts, each corresponding to an angle of incidence. The sub-pixels of each pinhole camera are labelled as  $r$ ,  $s$ , and  $t$ . Light passing through the right side, centre, or left side of the lens/pinhole will strike the  $r$ ,  $s$ , or  $t$  pixels respectively. In effect, each tiny pinhole camera forms an image of the main lens aperture, and this image captures the information about which subset of the light passed through a given sub-region of the main lens. If the object is in the plane of focus, as in Figure 2.7 (a), then all three of the pixels  $r$ ,  $s$ , and  $t$  of the centre pixel are illuminated. If the object is near or far, as in Figure 2.7 (b) and (c), then the light is distributed across the pixels in a manner that is diagnostic of depth. This makes the fundamental idea behind the microlens array-based cameras, where tiny pinholes are replaced with lenses of extremely small diameter arranged in compact size.

Figure 2.8 defines the LF camera, where  $z_0$  is the distances between the object plane and exit pupil, and  $z_1$  is the distance from the main lens to the MLA. As shown in the figure, the MLA is placed where the object is brought into focus by the main lens (imaging plane), with  $f_1$  being the focal length of the main lens, then assuming a thin lens, the lenses follow Equation 2.4.

$$\frac{1}{z_0} + \frac{1}{z_1} = \frac{1}{f_1} \quad 2.4$$

Also, it can be observed that the MLA is placed at a distance of  $z_2$  from the photosensor plane. The incoming light rays converging at the microlens will get diverged onto the photo sensors at different locations. To ensure sharpness of the image captured by the system, it is necessary to focus the MLA on to the main lens principal plane [21]. This is shown as the blue line in Figure 2.8. Since the microlenses are small compared to the main lens,  $z_1$  is at infinity. Hence the distance between the MLA and the photosensor can be approximated as Equation 2.5.

$$z_2 = f_2 \quad 2.5$$

Further, it is necessary to make sure that the light rays passing through the MLA are effectively utilised and recorded by the photosensor. Also, to avoid any leak or crosstalk of light rays with adjacent (neighbour) lenses other than the associated lens, it is important to match the diameter and distance values of the main lens and the MLA as given by Equation 2.6

$$\frac{d_1}{z_1} = \frac{d_2}{z_2}$$

2.6

This optical specification would result in a camera system that captures a filtered and sampled LF of an object on the photosensor. The sampling happens because there is a finite number of pixels under each microlens and the final image is due to the finite dimensions (2 planes and 4 dimensions, see Figure 2.3 ) [18]. Two perspectives are observed in this camera system. First, the region of the photosensor (pixel) under each microlens group (known as the microimage) and can be treated as a virtual pixel. Mathematically the microimage of the light field  $L(u, v, x, y)$  is obtained by the same  $(x, y)$  but with different  $(u, v)$ . Secondly, the group of light rays emerging from the same location of the main lens (shown as the blue line in Figure 2.8) but registered at different pixel locations. This can be referred as the sub-aperture image. Mathematically, the sub-aperture image of the light field  $L(u, v, x, y)$  has the same  $(u, v)$  but different  $(x, y)$ . The camera system emerging with the optical property is known as Plenoptic 1.0 and this model is considered for investigation in this thesis.

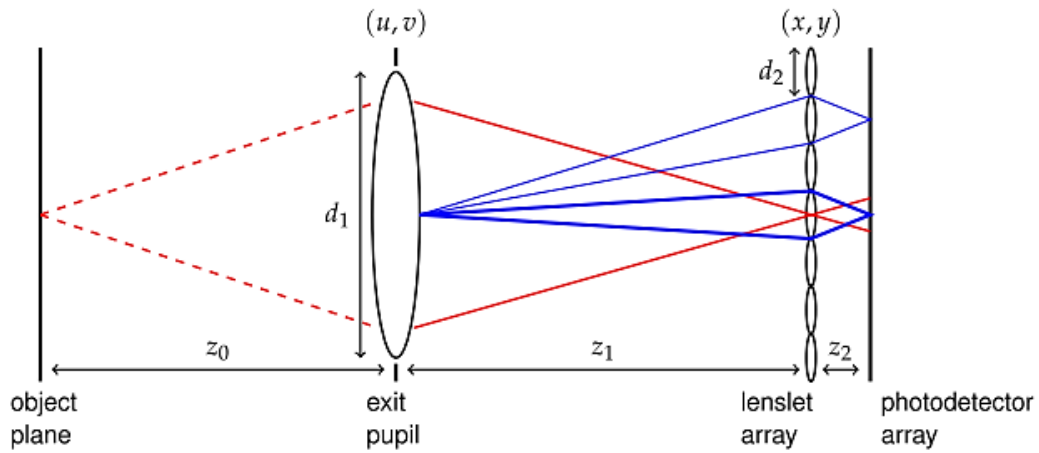


Figure 2.8 Location of lenses in the plenoptic camera system [18]

Plenoptic 1.0 cameras produce a final rendered image with very low resolution, typically one pixel per microlens. To tackle this problem, various methods have been developed, which are aimed to achieve better resolution from a plenoptic camera. Some captured more LF data by multiplexing many views, an idea explored in the programmable aperture camera [48]. This technique, however, requires a long acquisition time, and the object needs to be static. Combining an LF camera with a traditional camera is also a feasible solution [49].

A model of LF camera exists that overcomes the resolution obtained from the Plenoptic 1.0 model. The lens positions are shown in Figure 2.9 which is known as a focused plenoptic camera but Georgiev also called it the Plenoptic 2.0 camera [50]. In this camera system, the MLA is placed at a far distance from the main lens to focus on the image plane of the main lens rather than at the image plane as shown in Figure 2.8. This produces a relay system of reimaging the image at the image plane of the main lens. This indicates that photosensor has to be moved away from the main lens that satisfies Equation 2.7 with a focal length of the MLA,  $z_{1B}$  is the distance between the main lens image plane and the MLA, and  $z_2$  is the gap between MLA and the photosensor.

$$\frac{1}{z_{1B}} + \frac{1}{z_2} = \frac{1}{f_2} \quad 2.7$$

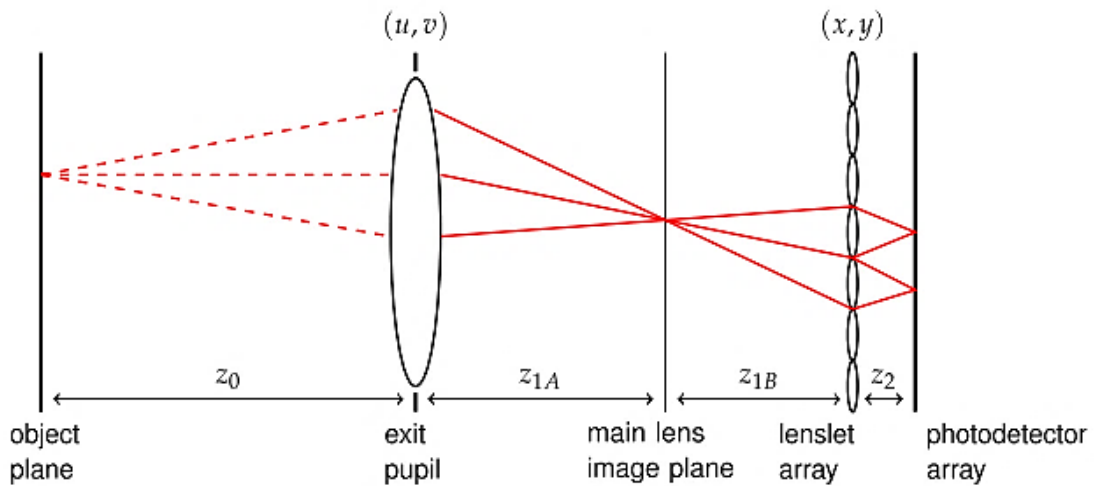


Figure 2.9 Location of lenses in focused plenoptic camera 2.0 system [18]

## 2.4. Commercially available MLA based light field cameras

Recent advancements in the field of image processing and computation have made it possible to use LF cameras for commercial and industrial applications. Also, another important reason for such a growth in the field of the LF camera is due to the advancement in manufacturing techniques in producing camera optics. Now it is possible to manufacture tiny optical lenses with high precision, which was not possible a few decades ago.

There are very few companies around the world manufacturing cameras based on LF technology. Lytro and Raytrix are well-known companies producing LF cameras based

on the microlens array technique. The Lytro Company is focused on manufacturing cameras for photography and entertainment and is comparatively cheap (around £ 100 to £ 1,000). The Lytro first generation (Figure 2.10a) and second generation Lytro Illum (Figure 2.10b) are the two variants manufactured by Lytro [51]. In comparison, Raytrix cameras (Figure 2.10c) have been developed for engineering applications. The Raytrix Company also produces a camera with special features depending on application [42]. There are many variants of cameras produced and they cost approximately £ 60,000 for the individual camera with additional cost for software and accessories [52] (see appendix 10.1 and 10.2 for camera specifications).



Figure 2.10 (a) Lytro first generation camera, (b) second generation Lytro Illum, (c) Raytrix camera [21, 22].

## 2.5. Conclusion

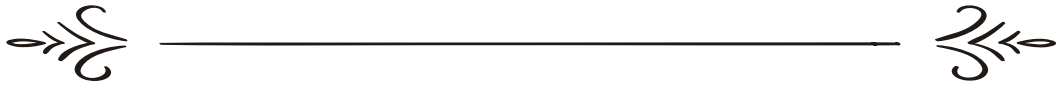
LF cameras record additional information along with the intensity information carried by the light rays. The optical elements used in LF cameras are very similar to conventional cameras with the MLA as the additional component, which enables the user to derive the directional information with suitable computational algorithms.



In this Chapter, the basic concept of LF acquisition has been described with appropriate diagrams with various optical elements used along with their positional details. The ray-space diagram helps to understand/ trace the light rays in 2D perspective. The concept of light rays passing through  $(u, x)$  and  $(x, y)$  planes define the basic concept of LF technology, and these planes are replaced by optical elements, the MLA and the photosensor respectively, in a LF camera. The plenoptic function,  $L(u, v, x, y)$ , will be used in upcoming Chapters for accessing information from the LF images.

Since the MLA plays an important role in providing directional data, it becomes necessary to understand the interaction of light rays with MLA. This Chapter introduced the methodology of LF imaging, i.e. ideas and factors necessary to generate the LF image, but no in-depth details are explained regarding the interaction of light rays with the MLA. Also, this Chapter defined the LF concept mathematically (2.2 and 2.3) with ray-space diagrams, while the spatial representation of LF would be more appropriate. The spatial representation of light rays, the interaction of light rays with the MLA and the photosensor, and the reason why LF technology is based on a pinhole camera model is explained in 3.

# 3



## LIGHT FIELD ACQUISITION

## *Overview*

A brief history of LF imaging and manufacturing advancements leading the way for new ways of capturing additional information using additional optical elements were explained in Chapter 2. Also, the interaction of light rays was shown with simple ray-space diagrams along with the positional details of lens elements for two different LF models with respect to the final resolution of the image.

However, greater detail of the MLA and its interaction with light rays must be identified to understand the LF imaging. The mathematical equations representing LF camera systems do not provide any information of light interaction with camera optical elements, hence light rays are re-traced within a single lenslet showing the working principle with the basic spatial domain is considered in this chapter.

Conventional cameras also record partial LFs, which is illustrated in this chapter by comparing with LF cameras. Also, the light ray models used to define and differentiate conventional and LF cameras are explained by considering point source objects, and the pinhole concept of LF technology is also demonstrated in this chapter.

This chapter helps to understand the construction similarities between conventional and LF cameras and features available for considering this technology for metrological applications are highlighted. Chapter 3 acts as a bridge, linking the conceptual LF model (Chapter 2) and practical LF models explained in Chapter 4 to 8.

## 3 Light Field Acquisition Using Microlens Array Based Cameras

### 3.1 Introduction

A camera is an instrument designed to record the radiance of light using optical elements and photosensitive film or a photosensor. The quality and resolution of the image depending on lens quality and dimensions of the photosensor respectively. The camera can be considered as a remote sensing device as it records/senses the object of interest without direct contact. The word camera comes from *camera obscura*, which means a *dark chamber* in Latin [53].

Modern cameras use a digital photosensor to record the light radiance and convert the information into digital values for further processing or storage. There are two main components used in digital cameras, the main lens and the photosensor, to capture the spatial information in  $X$  and  $Y$  directions (2D). In the field of metrology, conventional cameras (2D) are used as remote sensing devices to measure the height and width of an object with a fixed distance between the object and the camera. Many techniques have been proposed to capture depth information using 2D cameras such as; capturing multiple images of the same scene by changing the camera position ( $Z$  direction ) [2, 51–53], changing aperture diameter [57], and focal length. The method of changing camera position and focal length with respect to the object requires high precision components and takes multiple images for generating depth information in absolute units. The method of changing aperture diameter is suitable enough for some applications which require only non-metric depth data, i.e. different depths in the scene are represented by varying grey scale values and these values do not represent distance measured in absolute units (metres). Hence depth data generated by these techniques may not have a direct relationship to real-world measurements.

In the field of machine vision, depth information of a scene is highly useful, and many experiments have been performed to reduce the time taken to provide depth data using minimum sensors and hardware while maximising depth quality. Depth sensing instruments are finding many applications in the field of robotics and manufacturing industries, and new technologies are constantly being developed. Light field or Plenoptic cameras are one among the new methods that are capable of recording and

generating depth data. LF imaging has many interesting features available to offer the engineering community such as higher depth-of-focus, focus after capturing an image, depth information and a single camera operation. They also offer depth information of the scene in a single exposure which is a great advantage over many depths sensing instruments that require multiple images or multiple camera positions.

### 3.2 Image acquisition using conventional cameras (2D)

Conventional cameras are designed to capture the incoming light rays using a photosensor. These cameras can be considered as partial LF cameras since they capture only a part of the light radiance using a 2D photosensor. The new techniques available these days allows the recording of 4D or higher order information of the light rays. So, before understanding 4D LF imaging in detail, it is useful to recall the working principles of the standard camera.

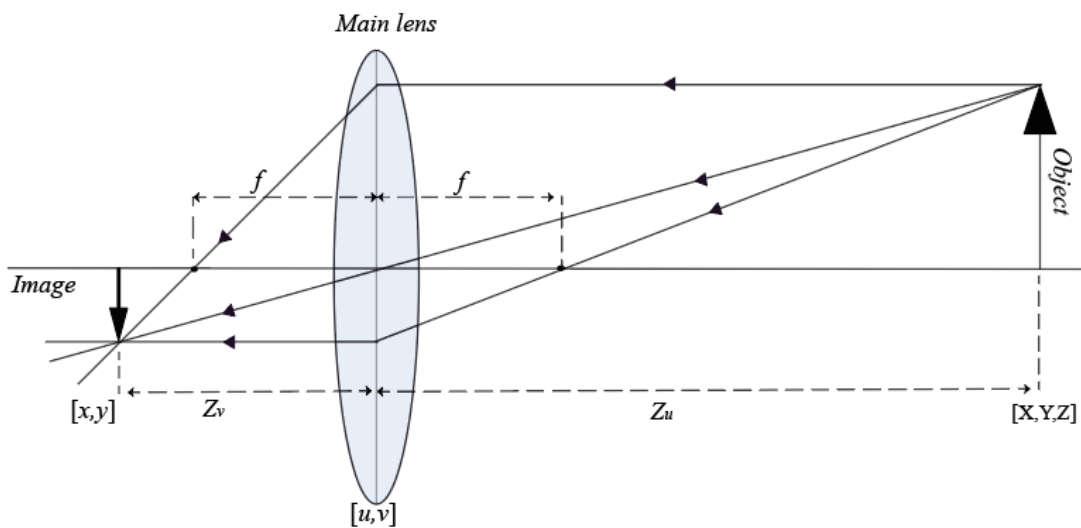


Figure 3.1 camera model (thin lens model)

The main lens and photosensor are the two important components of the basic camera model. In Figure 3.1, the light rays are bounced (reflected) from the object surface and travel in all possible direction and reach the main lens. All light rays in object space carry 3-dimensional information, i.e. positional data of the object ( $X, Y$ ) and depth data ( $Z$ ), with respect to the main lens of the cameras. The main lens inherits the property of the convex lens and hence tends to bend all incoming light rays from the object surface towards the optical axis of the lens. A sharp inverted image of the object is formed behind the focal length distance of the main lens known as the image plane. Since the light rays carry all information while passing through the main lens, it is denoted with

$[u, v]$ . By considering a 2D digital photosensor used to register the light rays,  $[x, y]$  represent the pixel location where the third dimensional data is lost. The distance between the lens and the image plane are measured by using the thin lens equation given by Equation 3.1, where  $z_u$  and  $z_v$  represent distance from the lens to the object and distance from the lens to image plane respectively, while  $f$  represents the focal length of the lens.

$$\frac{1}{z_u} + \frac{1}{z_v} = \frac{1}{f} \quad 3.1$$

For any given light ray, if the thin lens criterion is satisfied, the scene is in focus at the image plane. In a conventional camera, a light ray at some pixel location can be denoted with four variable functions  $(u, v, x, y)$  (see Chapter 2, sec 2.1), but in the practical scenario due to the lack of exact positional information of light rays travelling through the main lens  $(u, v)$  is generally neglected while representing an image, given by  $l(x, y)$ .

The captured final image has positional and intensity data of light rays travelling onto the camera sensor, but information related to distance travelled by the light rays or their direction is not recorded and hence no further data can be extracted from the final image. But depending on the amount of blur created by the final image, an approximate distance travelled by each light ray can be determined [58].

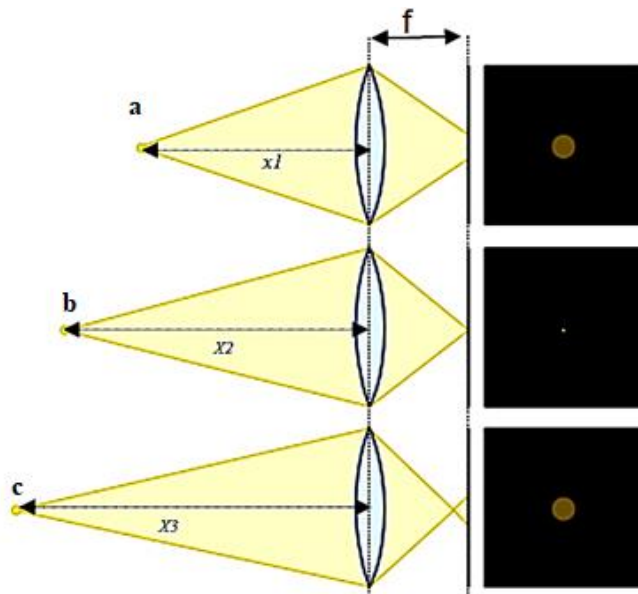


Figure 3.2 Blur-circle or Circle of confusion.

Figure 3.2 represents the formation of a blur circle depending on the distance travelled by light rays. The light rays from objects  $a$  and  $c$  travel distances  $x_1$  and  $x_3$  respectively before converging on the sensor forming blur circles. For object  $b$ , light rays travel distance  $x_2$  and form a sharp point image on the sensor after passing through the lens. Since, light rays for object  $b$  travel from a distance ( $x_2$ ) which is within the tolerance limit (converge to a single point) of the lens, a sharp image is formed at a distance  $f$  (on the sensor) and has a very small blur factor. Knowing the exact distance  $x_1, x_2$  and  $x_3$  with corresponding blur factor is difficult, however, an approximate distance can be calculated by comparing any other new blur factor with standard blur factor of objects  $a, b$  and  $c$  for the arrangement shown in Figure 3.2. This represents the additional information recorded by conventional cameras, but to measure distance with the help of blur factor, sequential images at a different distance from the object must be taken. Some research results have been published describing the depth of focus/defocus (blur) method, technical challenges and accuracy in measurement [56–59].

### 3.3 Image acquisition in MLA based light field cameras

LF cameras are similar to conventional cameras in functionality, i.e. to capture the radiance of light and register the information on a photosensor. But the optical behaviour of LF camera slightly deviates from conventional cameras due to the presence of the MLA. They are placed in between the main lens of the cameras and the photosensor as shown in Figure 3.3. This figure is not to scale because the MLA are drawn artificially large to make it possible to see them and the overall camera at the same scale. The MLA are very tiny compared to the main lens, and so is the gap between the MLA and the photosensor (see Chapter 4 for more details).

In Figure 3.3, the mechanism of capturing radiance is broadly classified into two spaces such as the object space and the camera space. In the object space, the light rays are reflected from the object surface in all possible directions to reach the main lens. In the camera space, these light rays further interact with two lenses (main lens + MLA) to register on the photosensor. The region of each lens is highlighted with different colours to indicate the boundary of their effect on the light rays. The main lens bends the incoming light rays towards the optical centre as a function of the focal length, where all light rays are bundled with respect to a single main lens.

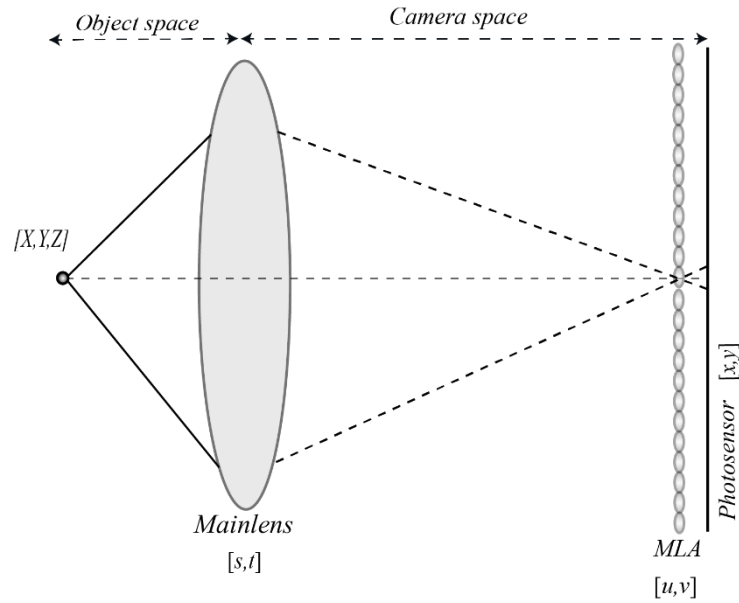


Figure 3.3 Effective boundary region classified according to the position of lenses

Further, the light rays pass the effective boundary region of the main lens and interact with the MLA. The MLA inherits the same optical properties as the convex lens, the light rays interacting with the MLA bend towards the optical centre. Since the photosensor is placed at focal length distance from the MLA, the light rays are registered as a function of direction and location of parent MLA. Hence, the pixels under the microlens unit (pixels of the microimage) represent intensity values of a point object (object surface) in a different direction which is equal to the number of pixels in the microimage. In this arrangement, the microlens plane is the imaging plane, and the size of the microimage is the spatial sampling resolution of MLA lenslet (see Chapter 2, sec 2.3).

### 3.3.1 Pinhole approximation of LF cameras

To better understand the interaction of light rays with the MLA, consider the two-point object  $a$  and  $b$  as shown in Figure 3.4, where point source objects reflecting light in all possible directions pass through the microlens optical centre  $o$  to form a point image  $A'$  and  $B'$  respectively [47, 63]. The lenslet of the MLA shown in Figure 3.4 consists similar optical properties as the main lens, but major distinctions are with focal length and the size of lenslet (diameter) (see Chapter 4).



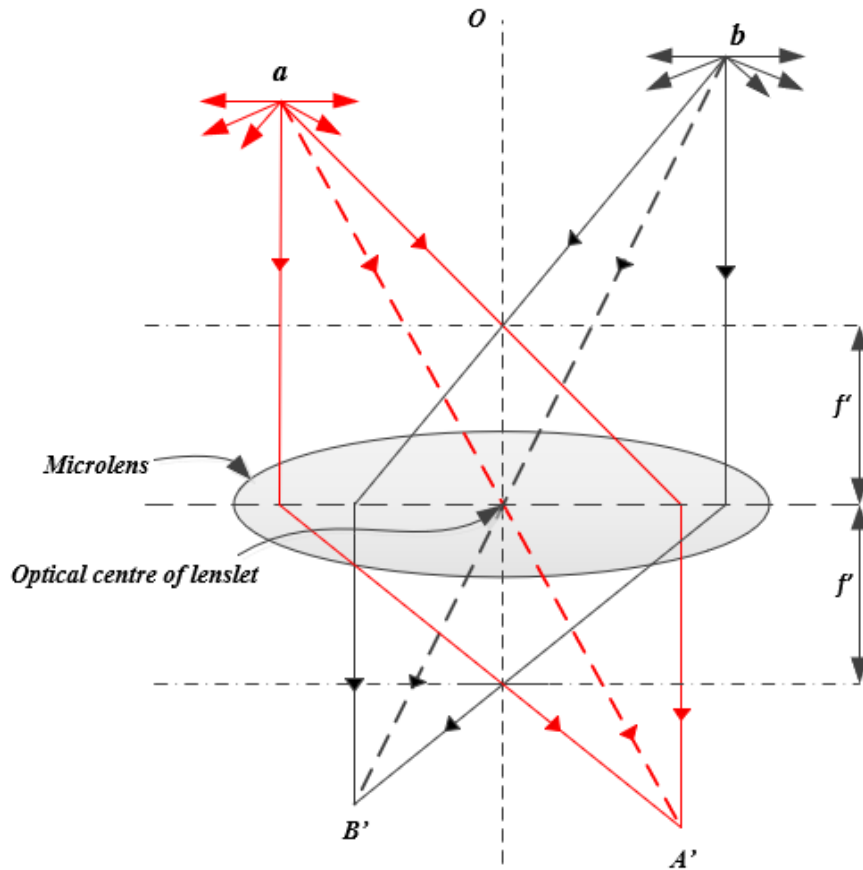


Figure 3.4 Light propagation through the microlens

The main lens is not shown in Figure 3.4 since the light rays carry directional and intensity information in between the main lens and the MLA, similar to a conventional camera. The directional data will be extracted from the light rays passing through the MLA. In Figure 3.4, all lenslets inherit the same optical property and hence a single lenslet is shown with a two-point source.

Three rays from each point object have been considered among an infinite number of rays that travel through the lens and meet at the image plane. The incoming light rays can be divided into two models depending on the entry and exit of a light ray from the lens. The light rays passing through the optical centre of the lens are primary or chief rays (Figure 3.4, dotted line and Figure 3.5 (top)) and the rest are secondary rays (Figure 3.4, thick line and Figure 3.5 bottom). Since the point objects are at different distances from the lenslet, the images of the point source are also formed at different distances behind the lenslet.

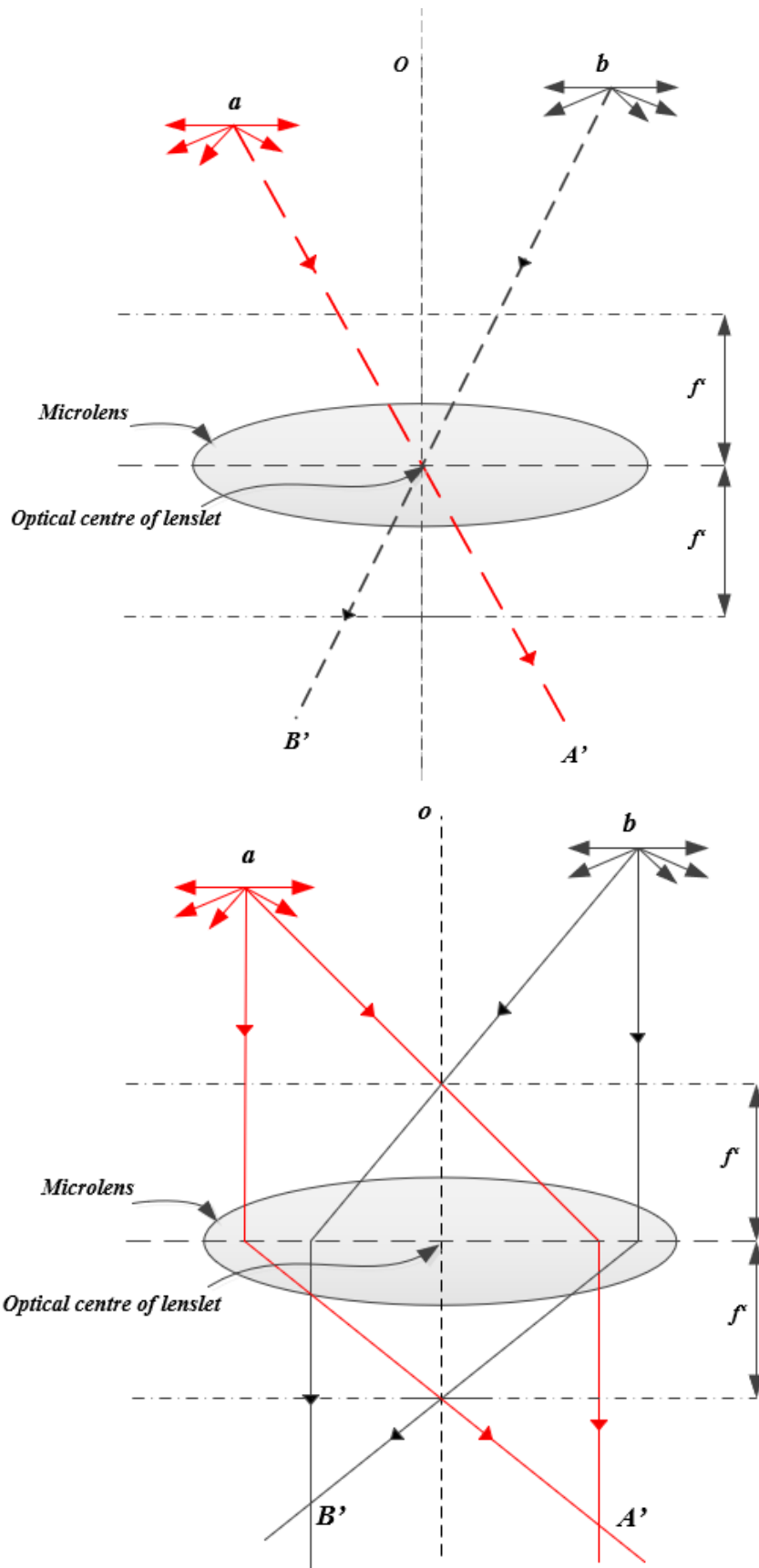


Figure 3.5 Primary (top) and secondary (bottom) light ray models

If secondary rays are considered to generate a point image of A and B, the intensity of the point image may be low because the point image is formed by summing two light rays  $A$  and  $B$  rather than three light rays. However, the information gathered at the image plane remains the same if the intensity term is neglected. The same holds true if only primary light rays are used to generate a point image at the image plane. It can be observed that information of both point objects ( $a$  and  $b$ ) are available at any distance from the lenslet in a primary ray model. In contrast, with the secondary ray model, the information is only available at the imaging plane and diminishes moving away from the image plane since the light rays move further apart and interact with other light rays in the practical scenario.

Since most conventional cameras follow the secondary ray model (see Figure 3.5), information is lost in the case where the photosensor is not aligned with the image plane of the lens and results in blur image. In contrast, the primary ray model retains the same data at different distances from the lens and this model is also known as the pinhole model. This is the basic concept of LF imaging and post-focusing, where different primary and secondary rays are summed at various distances from the microlens to regenerate originally lost information computationally with the assumption that light rays travel in a straight light.

### **3.3.2 Light propagation through MLA**

The LF camera works on the basic principle of the pinhole camera model [47, 63], where each microlens in the MLA acts as a small individual camera, recording a small section of radiance using the pixels under each microlens, known as the microimage. Every microlens group is approximated to record light rays passing through the microlens optical centre, acting as a pinhole, from different viewpoints and hence the LF image is a set of different views of a scene.

Figure 3.6 represents two-camera configuration with slight variation in their respective internal optical design (camera lens system), that captures light rays from the same object. The object surface reflects light rays in all possible directions represented by  $a'$  to  $f'$ . All the light rays carry the same information but travel in slightly varied angles with respect to each other. The main lens bends the incoming light rays towards the optical axis of the lens at the point of interaction. By placing a photosensor at the focal distance, a point information of the object surface is generated (see Figure 3.2 for

further details). In contrast, if an MLA is placed at the focal length distance of the main lens and by approximating the MLA to behave as pinhole model, the light rays pass through the pinhole and get registered on the microimage pixels with respect to the angle at which they were reflected from the object surface (also depends upon the numerical aperture (NA) of the main lens). The microimage consists of point information of the object surface in terms of different angles or different views. The order of light rays is flipped under Plenoptic 1.0 after passing the MLA plane, but with multiple lens systems in modern cameras, the light rays can be flipped before reaching MLA plane and hence resulting the light rays in the same order as they were created. It can be seen in the Figure 3.6 (left) that the light rays are registered in the same sequence as they were created, where as in Figure 3.6 (right) the information registered in opposite sequence illustrating the role of modern optics/lens system used in the camera design.

In Figure 3.6, even though only seven primary light rays are shown to be passing through the optical centre  $o$  at the MLA plane  $u$ , there would be contributions from many primary light rays to form pixel data in the real scenario. So, by knowing pixel size, microimage resolution, the optical centre of microlens ( $o$ ) in terms of pixel coordinate  $(x, y)$  and the distance between MLA and the photosensor, the angular direction of all the light rays recorded on the microimage pixels for a given microlens can be calculated accurately. Also, light rays can be traced beyond physical photosensor distance to the MLA.

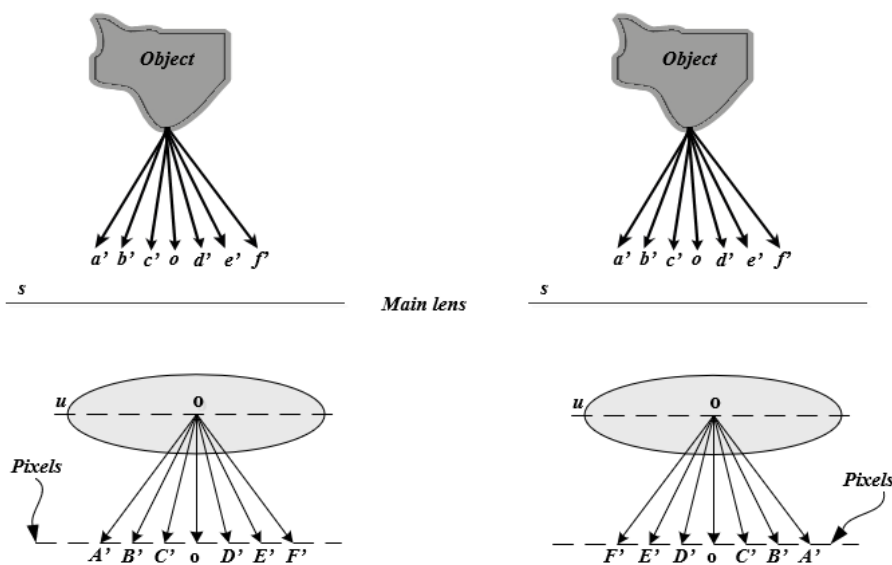


Figure 3.6 Views of an object recorded by a single microlens illustrated using two cameras

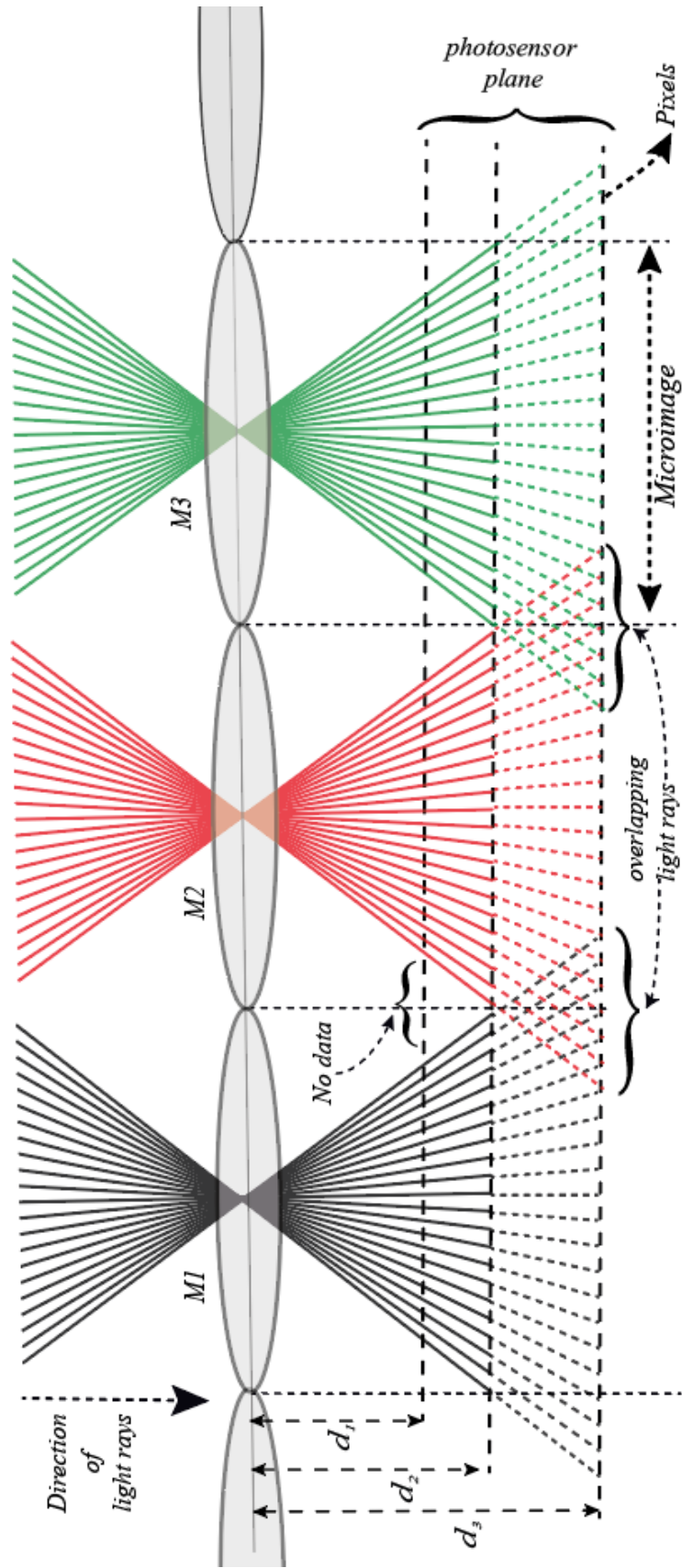


Figure 3.7 representation of distance-dependent microimage resolution and overlapping of light rays

The quality of the image and number of light rays recorded by each group of the microlens depends on the distance between the MLA and the photosensor. The distance should be precisely adjusted so that all pixels of the microimage receive intensity information from incoming light rays. In the illustration shown in Figure 3.7,  $d_1$ ,  $d_2$  and  $d_3$  are the distances between the photosensor plane from the MLA. The distance  $d_2$  in Figure 3.7 represents the acceptable distance between the photosensor and the MLA to fill all microimage pixels with intensity data.

In comparison with  $d_2$ ,  $d_1$  and  $d_3$  distances pose problems for recording perfect LF data. While considering  $d_1$  as the distance between the MLA and the photosensor, some of the pixels under the microlens are not hit by light rays thereby no information is recorded (marked as ‘No data’ shown in Figure 3.7). Also, overlapping occurs within the microimage pixels since many light rays are gathered on a limited number of pixels. In the case of the  $d_3$  distance, the light rays extend into the neighbouring microimages and hence overlapping of data occurs at corners of each microlens group (if the equation 2.6 is not satisfied, the cross talk between neighbouring microlenses are known as overlapping of data). The light rays from the  $M2$  microlens get recorded on the  $M1$  microlens and  $M3$  microlens (marked as ‘overlapping light rays’ in Figure 3.7 ). This shifting of light rays due to distance  $d_3$  pose a problem during post-focusing features and results in pixelated images. Since the LF model has been approximated to a pinhole camera [47, 63], the focal distance of the MLA can be considered as the approximate distance between the MLA and photosensor. Using the gap between the MLA and the photosensor, the resolution of the microimage can be varied (physically changing the gap size), while the pixel size used in the photosensors can be varied to achieve higher microimage resolution depending upon user requirement.

### **3.4 Features of MLA based light field cameras**

The additional data gathered by the LF cameras based on the MLA tend to exhibit many features that are not seen with conventional cameras. Conventional cameras record the light rays using the secondary model (see sec 3.3.1) and hence the final best-focused image is a resultant of the average between many light rays. In contrast, the intensity recorded behind the MLA can be further classified based on the direction of light rays. Such classified images are called sub-aperture images that represent 2D camera views of the scene. The additional angular data also facilitates the process of de-blurring the

image or a part of the scene after capturing the image, known as post-focusing or focus sweeping. This is one of the primary advantages of MLA based LF cameras compared to conventional cameras.

### 3.4.1 Sub-aperture images

A sub-aperture is a collection of pixels of particular direction (pixel location in the microimage) to generate a 2D image of the scene in front of the camera, as defined in Figure 3.8. However, the final raw image of the LF camera is different when compared to data recorded by a normal 2D camera. In an LF image, each pixel is represented by four positional variables  $(u, v, x, y)$ , where  $u$  and  $v$  represent the location of lenslet in the MLA (in terms of pixels), while  $x$  and  $y$  are the positional variables of the photosensor (in terms of pixels). Since, the final image can be represented by more than four positional variables, the LF image inherits the feature of transforming the 4D image into 2D sub-aperture images also known as perspective shift images (see Chapter 2, sec 2.3). The perspective shift images can be described as capturing light rays of point object passing through the same aperture location but registered under different lenslets. Mathematically, these images are the LF  $L(u, v, x, y)$  with the same  $(u, v)$  but different  $(x, y)$ . Different combinations of positional variables result in different perspective images, that can be used depending upon application requirement, thus the same point of the object can be viewed from different angles. In the context of this description, a microimage is a group of pixels, and the microlenses group is a group is individual microlenslets.

Every individual pixel of a microlens receives light rays travelling from a different direction which are reflected from the same object location. Hence data recorded by each MLA group represents directional light rays from a point source. Figure 3.8 represents LF data recorded on a 2D sensor, where  $q_x, q_y$  represents the positional variable of the MLA (direction of light rays) and  $p_u, p_v$  represents the positional variable of the photosensor. The resolution of microimage in Figure 3.8 is 16 (4 x 4 pixels) and each pixel represents data of the light rays recorded on the photosensor from different directions. Hence individual microlens groups record 16 different angular information and thereby allowing the perspective shift feature in the LF cameras. For example, the top left corner pixels are selected from every microlens group, which are rearranged to represent a final image (sub-aperture image).

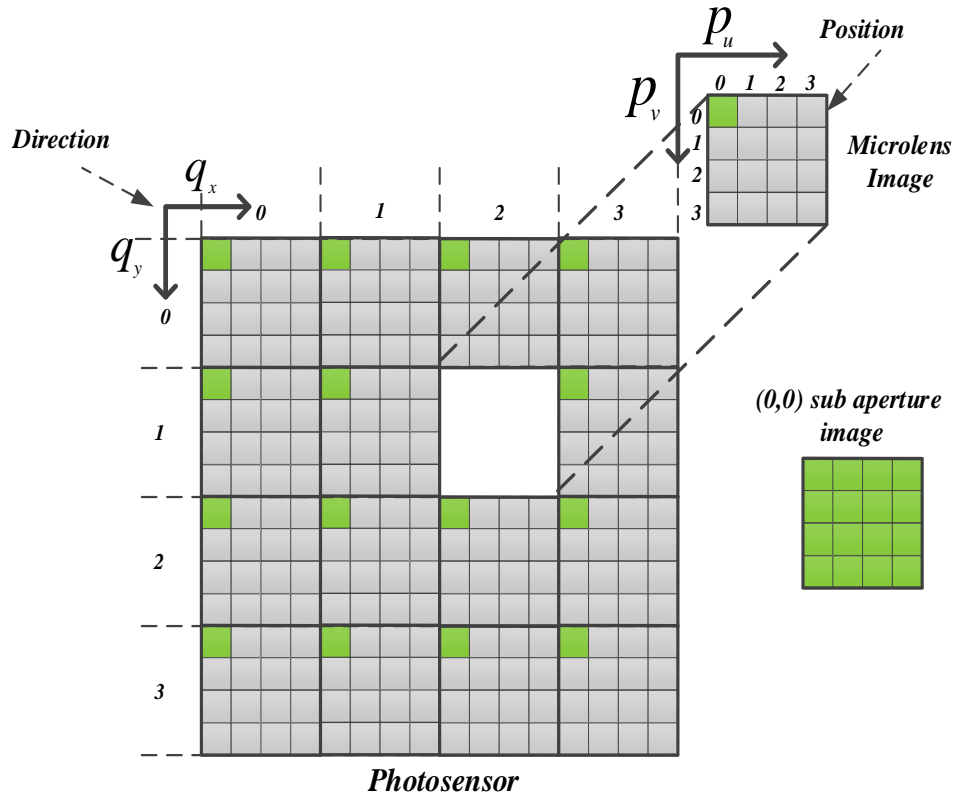


Figure 3.8 Light field image on a 2D photosensor

This perspective shift image ( $[0,0]$  camera view) can be represented by  $(q(0:3,0:3), p(0,0))$ . By selecting different pixels from each microlens group, perspective images with different angular values of the same light ray are assembled in the final image.

The microimage resolution of the Lytro-I generation camera is 100 pixels and hence 100 individual light rays are recorded, which represent 100 angular information, i.e. a point in the scene is captured from 100 different view, where every view differs slightly with respect to its neighbouring view. An image of an engine block recorded using a Lytro-I generation camera is shown in Figure 3.9 (top). By selecting a single light ray from the MLA and corresponding centre pixel from the photosensor, a perpendicular view of the scene is generated and represented in Figure 3.9 (bottom). There are 100 complete images of the scene with different angular information represented by an individual image. By selecting appropriate pixels and directional light rays within a suitable algorithm, a smooth shifting of views can be achieved.



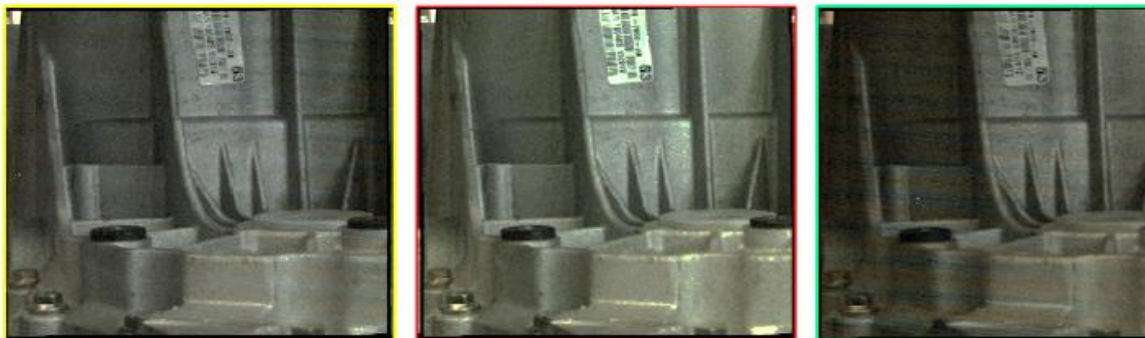
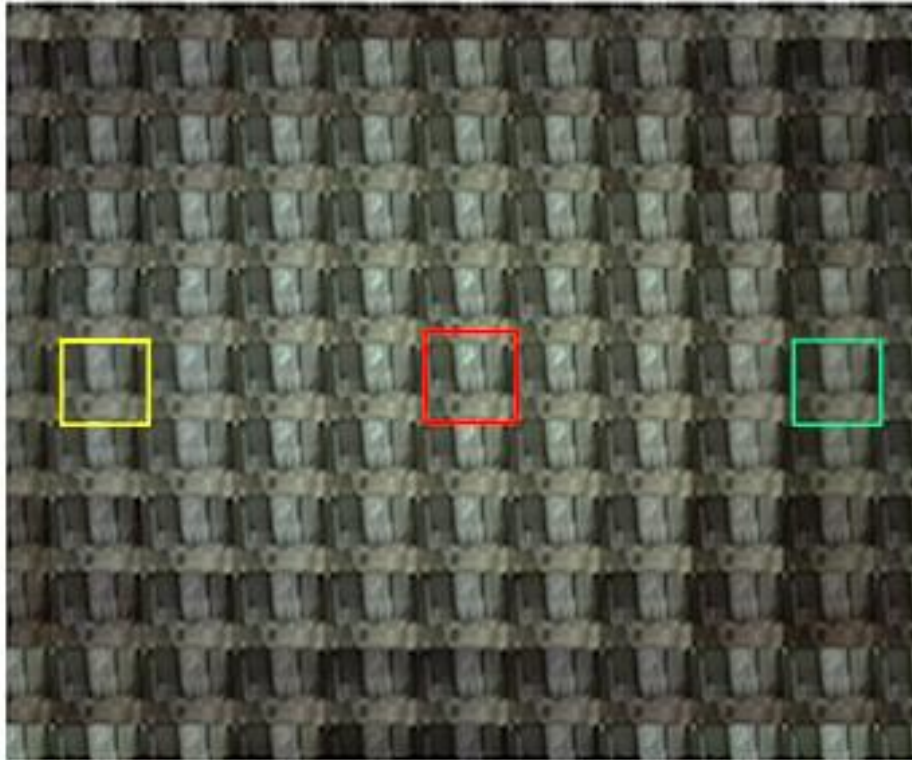


Figure 3.9 2D camera views generated by Lytro Desktop software tool (top) and enlarged corresponding sub-aperture views (bottom)

### 3.4.2 Post-focusing

The post-focusing feature is one of the interesting features, that enables to reprocess the blurred regions of an image. Post-focusing is a technique of changing the focus over the image plane where objects in a scene are brought in/out-of-focus depending on the distance from the camera. The basic principle for post-focusing has been mathematically explained in [13, 20].

Three microlenses of the same focal length in an MLA are shown in Figure 3.11, where the photosensor is initially placed at a distance  $d'$  from the MLA. In this situation, due to the pinhole approximation, light rays pass through the microlens optical centre and

are recorded on pixels in each microimage group. In this configuration, it is noticeable that there is no overlapping of light rays at the corner of each microimage. At this stage, objects far away from the camera are in focus and the rest are out of focus. The light rays recorded by the photosensor pixels at the distance  $d$  are highlighted with a circle for the microlens  $M2$ . The light ray recorded by the pixel under consideration is denoted as  $p_i$  and this pixel is next to the pixel that records the light ray passes along with the optical centre of the MLA.

If the photosensor is physically placed at a distance of  $d''$  instead of  $d'$ , then the light rays would travel a longer distance of  $d''$  and get recorded on the photosensor. In this situation, there will be overlapping of light rays happening on pixels of each microimage group. For example, the pixel recording the light ray from the optical centre of  $M2$  at a distance  $d''$  get additional light rays from microlenses  $M1$  and  $M3$ . A similar phenomenon happens with the rest the pixels in a microimage and in all microlens groups. The pixel under consideration ( $p'_i$ ) receives data from microlens  $M2$  and  $M1$ , and this data is recorded on a different pixel when compared to the same scenario with distance  $d'$ . Also, the data recorded by  $p'_i$  pixel is different with respect to  $p_i$  pixel and this is because the photosensor was moved to the new distance. In this situation, some part of the scene will be in focus compared to the initial raw image of the LF.

Since the microlenses are considered to behave as secondary light ray model (Sec 3.3.1), the light rays are approximated and re-traced to travel in a straight line after passing through the MLA and this is an important assumption for post-focusing features in this section. If the photosensor physical distance  $d'$  is retained and pixel  $p_i$  is retraced further towards distance  $d''$  virtually, then we end up reaching the pixel  $p'_i$  since the light ray travels in straight line. If the same technique is followed for all the pixels, the resulting data matches with the data of the photosensor physically placed at  $d''$ . To perform such virtual movements of the photosensors, fewer details about the MLA are necessary such as the microlens centre location with respect to the photosensor pixel. This information can be calculated manually [64] or supplied from the MLA manufacturer.

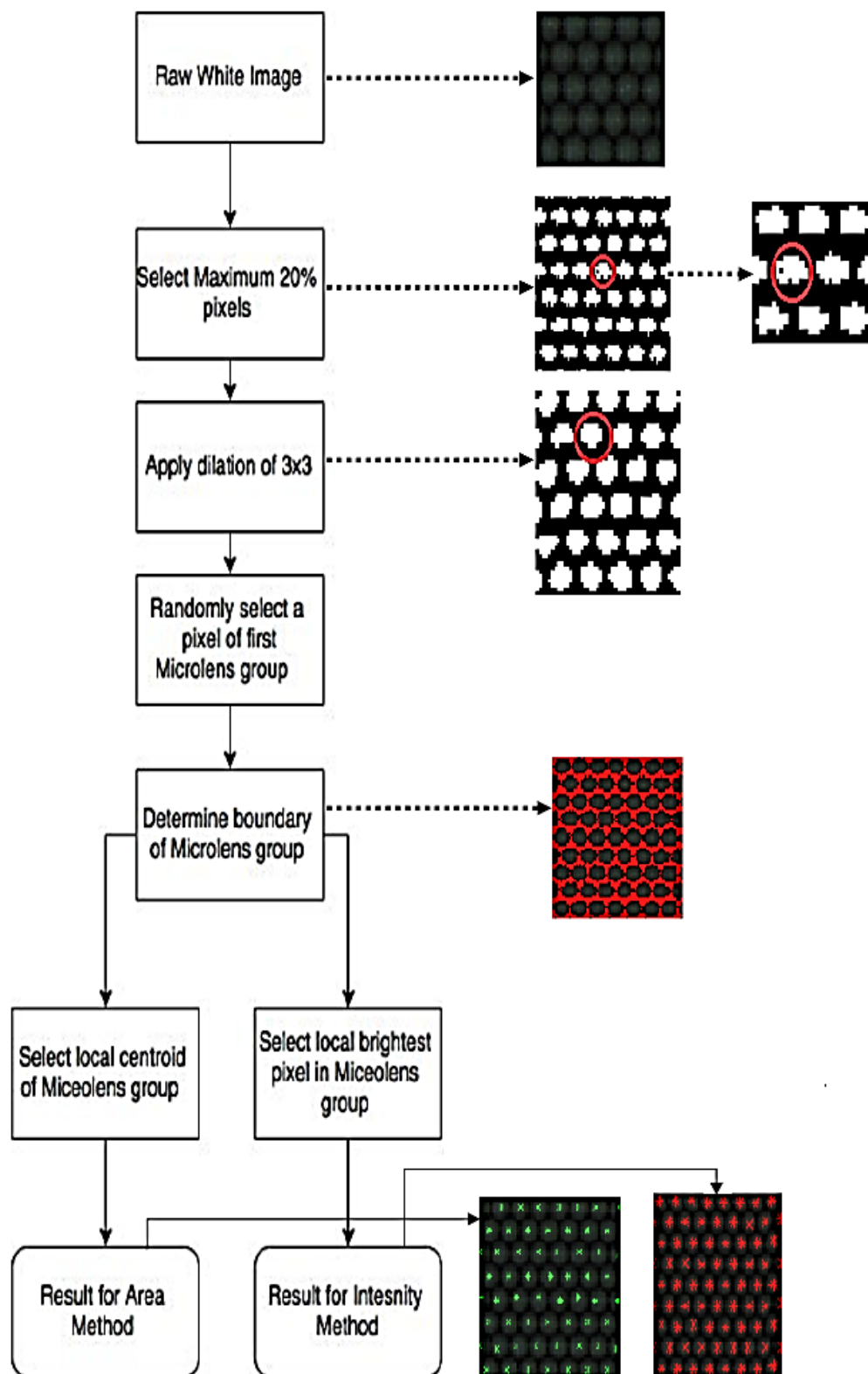


Figure 3.10 Manual procedure to calculate microlens cord manually

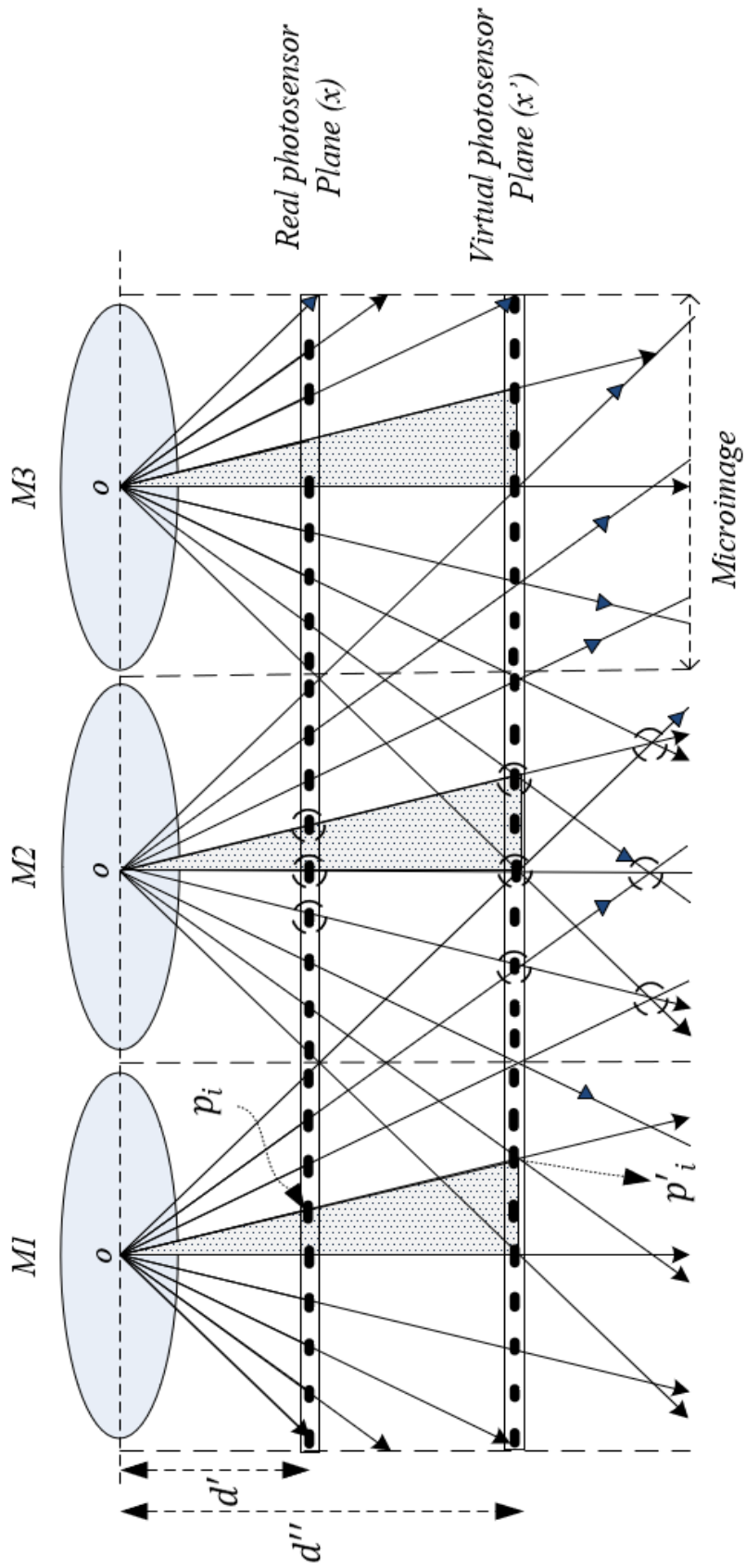


Figure 3.11 Real and Virtual photosensor positions

To achieve this movement of the photosensor virtually, the location of the pinhole centre of every microlens must be collected with respect to the photosensor pixel coordinate system, known as the MLA cord. An MLA cord is the collection of optical centres of every microlens in the MLA in terms of pixel coordinates. This information is provided by the manufacturer designing the MLA-photosensor assembly or calculated manually using white images (see Chapter 4, sec 4.4).

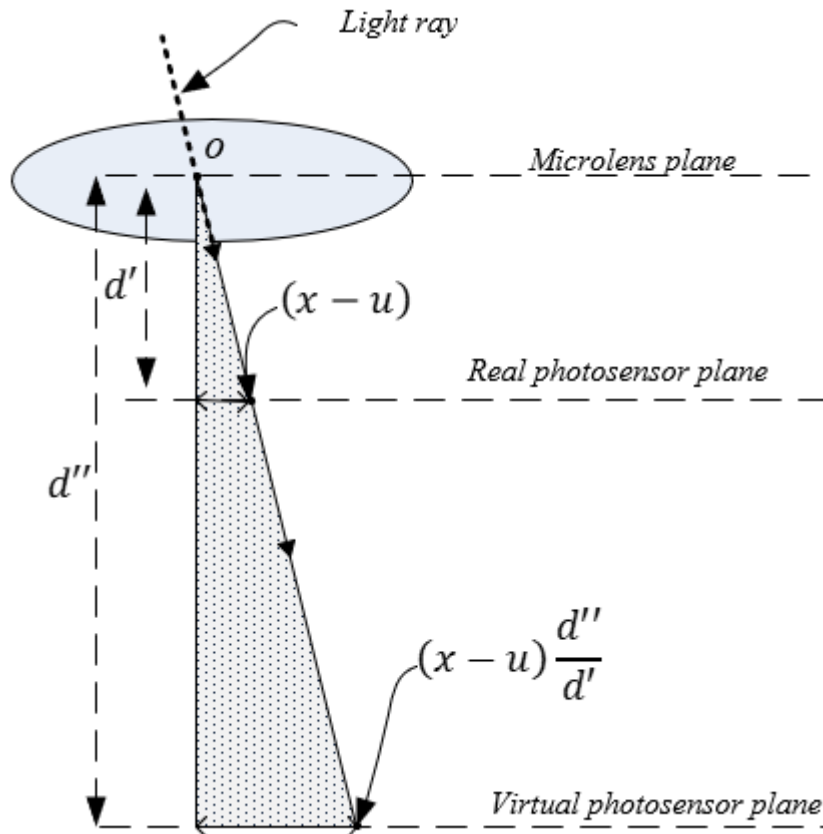


Figure 3.12 Ray-space coordinates

The ray-space coordinate diagram (Figure 3.12) is a geometric reconstruction that illustrates how a ray is parameterized by considering physical pixels and MLA location to generate virtual pixel data and hence an entire LF' image (virtual LF data). This can be re-parameterized by the intersection of the light ray with the MLA and photosensor planes. In Figure 3.12,  $u$  is the microlens centre,  $x$  is the pixel location on the real photosensor under consideration with the parent MLA location of  $u$ , so the pixel under consideration is  $(x - u)$  and  $x'$  is the resulting location of the light ray on the virtual photosensor. By similar triangles, the illustrated ray that intersects the lens at  $u$  and the film plane at  $x$  also intersects the  $x'$  plane given by Equation 3.2, for a single pixel.

$$x' = \left( u + (x - u) \frac{d''}{d'} \right) \quad 3.2$$

The diagram only shows the two-dimensional light rays involving the  $x$  and  $u$  planes, but in the three-dimensional scenario, there will be additional  $y$  and  $v$  planes forming complete 4-dimensional LF data. As a result,  $\alpha$  is defined as ( $\alpha = \frac{d'}{d''}$ ) which is the relative depth of the film plane and the recorded LF is given as  $L_F(x, y, u, v)$ . Equation 3.3 and 3.4 (for the whole image) represents the virtual 4D LF data,  $L'_F(x', y', u, v)$ , that results from focusing at different depths corresponding to the  $\alpha$  value.

$$L'_F(x', y', u, v) = L_F \left( u + \frac{x - u}{\alpha}, v + \frac{y - v}{\alpha}, u, v \right) \quad 3.3$$

$$L'_F(x', y', u, v) = L_F \left( u \left( 1 - \frac{1}{\alpha} \right) + \frac{x}{\alpha}, v \left( 1 - \frac{1}{\alpha} \right) + \frac{y}{\alpha}, u, v \right) \quad 3.4$$

### 3.4.3 All-in-focus and depth data

The images captured using conventional cameras tend to generate blur regions when the object does not fall in the depth-of-field of the camera lens system (Sec 3.2). In some applications, the telecentric lens system is used to eliminate unwanted blur and maintain constant field-of-view [65], especially in machine vision applications. This requires additional optical arrangement in case the camera used was not supported with the telecentric lens system.

LF cameras offer the post-focusing feature, which enables the user to focus the blur regions after capturing an image. These post-focused images are called image stacks which vary in sharpness and blur regions. In this context, the word sharpness can be defined as the variation of the pixel with respect to its neighbouring pixels. By calculating sharpness (see Sec 8.3, Figure 8.5) variations in every pixel throughout the image stack, it is possible to calculate the stack count per pixel corresponding to the highest sharpness variation. The resulting image will be a blur-free image that generates a similar effect as a telecentric lens system. This stack count corresponding to the focused pixel serves as the relative depth map which is an additional result available with the LF cameras.

In Figure 3.13, the all-in-focus feature available with the Lytro family of cameras is demonstrated along with post-focusing and a relative depth map. The Lytro company provides the Lytro Desktop Software (LDS) to generate different results from the LF data captured using their cameras. It can be noticed that the top row, three images (left to right) are focused at different distances from the camera (far to near). The bottom row consists of an all-in-focus image (left) and relative depth map (right).

Advanced computational photography has made it possible to refocus at different depths after light rays are recorded by the photosensor in a single exposure. Figure 3.13 shows three images taken by a Lytro-I generation camera with automatic exposure. Each image represents a focus at different depths from the camera. The top left image is focused on the green/blue colour object/gear approximately 400 mm away from the camera, the top middle image is focused black colour object at 200 mm from the camera, and the top right image is focused on the yellow colour object/gear placed close to the camera at a distance of 10 mm from the camera. While the bottom left image represents the all-in-focus image and bottom right image corresponding depth map generated by the Lytro-I generation camera with the LDS.

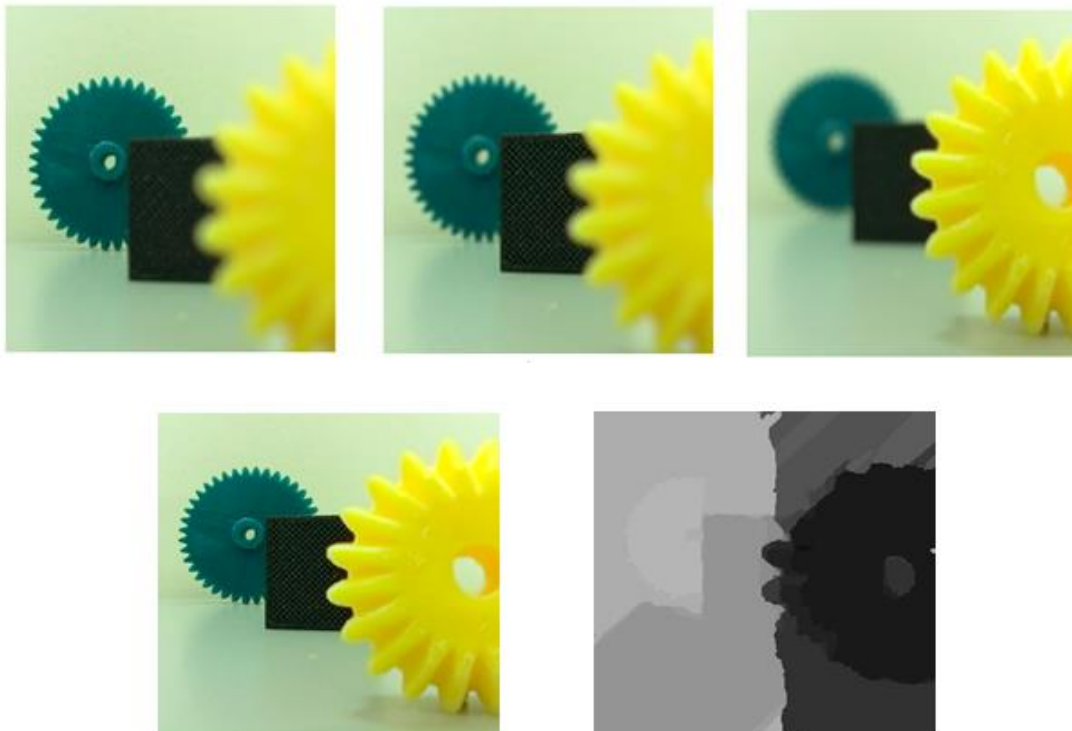


Figure 3.13 Features of light field technology available with the Lytro family of cameras

### **3.5 Potential features of MLA based light field cameras for metrology**

Machine vision plays a crucial role in many engineering applications and it is achieved using different methods depending upon the nature of the application. In machine vision applications, real-world data is captured using sensors and interpreted later by computer programs to generate information about the surrounding environment. In some applications, two-dimensional data is sufficient and digital cameras are widely used to capture data.

But in some applications, the depth information is crucial along with two-dimensional data for which multiple sensors are used. Industrial metrology is one of many engineering fields where 3D information is necessary along with the precision of measurement. Conventional cameras pose many challenges to capture 3D information under constraints such as minimal sensors, high mobility (flexibility of carrying instrument) and easy hardware set-up.

Further, to generate depth information using conventional cameras, moving the camera relative to the target is necessary, while stereo vision requires a detailed understanding of relative positions of cameras for better results. These techniques become very complex and time expensive since experimental setup must be reorganised once any component or place of the experiment is changed. In the field of industrial metrology, conventional camera techniques are widely used with complex systems to generate additional information along with the 2D image. One such arrangement involves registering accurate  $Z$  direction movement of the camera relative to object to generate 2D slices of 3D data and redesigning stereo vision system by mounting two fixed cameras to a single PCB, thereby introducing mobility feature to the system.

The MLA based LF cameras are very compact in design with high mobility. LF cameras use MLA optical elements to capture additional information and can be mounted within the camera body, hence fewer moving parts. The LF cameras have similar dimensions to mass ratio, thereby flexible to move without the need to change any camera settings. Moreover, LF cameras operate with a single photosensor which prevents additional calibration required in stereo vision cameras. Also, LF cameras provide additional features such as post-focusing, all-in-focus image and depth data.



Hence, LF technology exhibit promising methods to be implemented in metrological applications.

In the last few years, few noticeable developments have been made in implementing LF methods in metrological applications [62, 63], however, it is still at the initial stage compared to other existing methods. Since LF cameras depend on the light rays reflected by the object surface and do not project any sort of light on the object, they can be categorised under non-contact, passive method of data collecting devices, as shown in Figure 3.14.

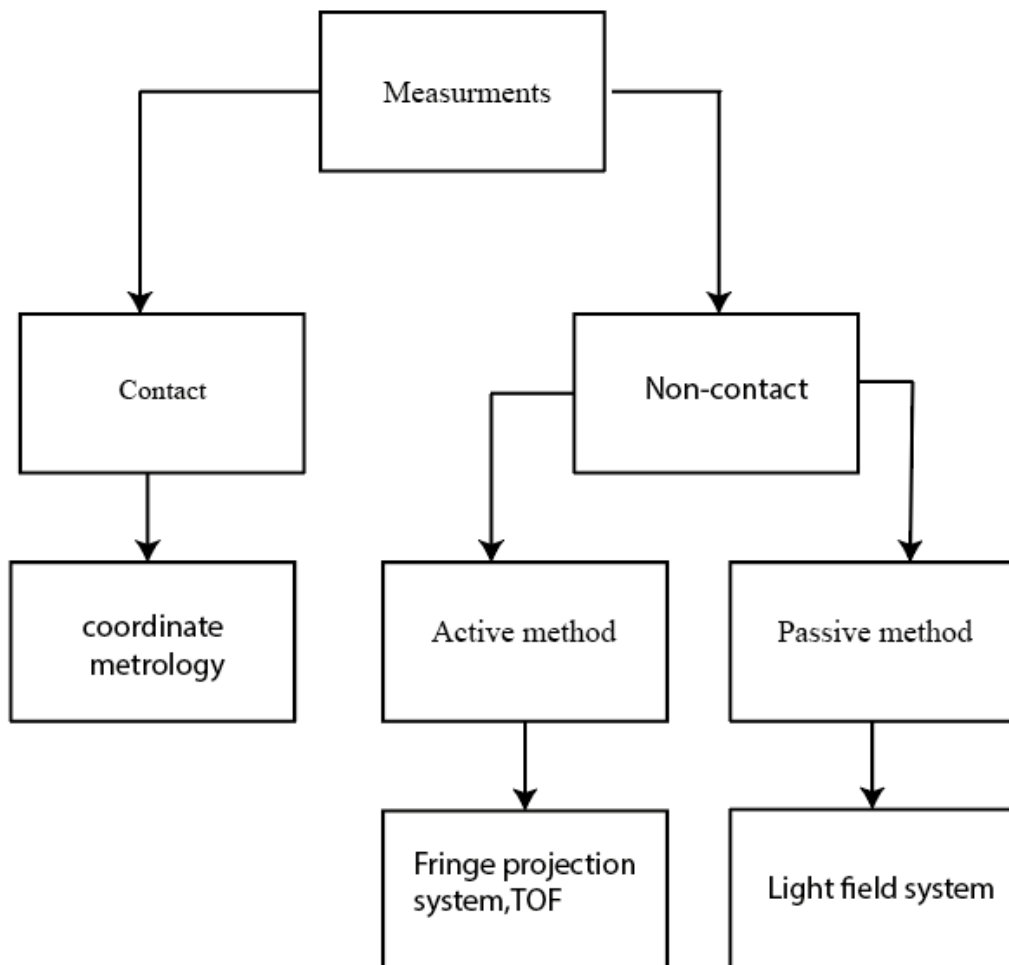


Figure 3.14 Classification of measuring devices

### 3.6 Conclusion

LF cameras inherit many interesting features that show promising results and potentially, may be suitable for many engineering applications. This chapter has mainly focused on illustrating the principle behind these interesting features exhibited by LF

cameras and features that can be used in engineering applications, especially targeting industrial metrology.

The common features and optical elements between conventional and LF cameras, as illustrated in this chapter, help us to understand the simplicity behind the new technology and rethink the areas where LF camera can be used for better results. Further, it is difficult to understand the interaction between light rays and optics only by using ray-space diagrams. Hence, simple ray tracing techniques were used in this chapter to illustrate the light interaction with the MLA lenslet. The ray tracing has been done based on a pinhole camera model which the LF cameras exhibit, and the reason behind this assumption has been described in this chapter.

Many important features emerge by defining how LF cameras work on the basic principle of the pinhole camera model, and the important reason why conventional cameras fail to show these features can be noticed in this chapter by considering primary and secondary light ray model.

Since, the Lytro cameras are the primary LF capturing devices used in this work, some of the features and corresponding results from these cameras (Lytro-I generation + Illum) are shown. These results are explored and developed in the next chapters, with chapter 4 generating additional details required about the special lens system (MLA).



# 4



## BASIC PARAMETERS OF MICROLENS ARRAY

## *Overview*

The microlens array is one of the key elements of the optical components used in an LF camera, that enables recording of intensity along with the direction of incoming light rays on a photosensor. It is important to understand the role and behaviour of the MLA so that a camera can be designed to incorporate all features available with an LF camera. The MLA parameters allow the user to understand the interaction of light rays with the MLA and hence ease the process of designing algorithms to use the 4D data gathered by LF camera. Some of the key parameters of the MLA are the focal length and the number of microlenses. Additional properties accessed after mounting on the photosensor are the microlens pitch size and the MLA cord (see Sec 3.4.2). In addition, some misalignment issues can be introduced while mounting the MLA and corresponding calibration parameters are calculated using the MLA parameters.

Existing research results provide information about finding some of the MLA parameters with the assumption that the user has already gained fundamental information (for example, pitch size) of the MLA, provided by the MLA manufacturer or via manual calculation by the user. For applications where an LF camera is assembled with components manufactured from different companies, the key parameters may vary once the MLA is mounted on the photosensor, for example, the microimage and hence the MLA pitch count in terms of photosensor pixel. The procedure to obtain all parameters of MLA without the need of any pre-calculated data is explained in this chapter.

The necessary mounting issues are discussed along with correction measures where applicable, along with corresponding ray tracing diagrams. The effect of misalignment on MLA cord detection is explained with the help of light directed MLA cord detection method. The major difference between the light directed method and other existing methods are briefly described and MLA cord generated using these methods results in a difference of 2 to 3 pixels. The light directed method is incorporated in the cord detection algorithm and is explained in this chapter using the raw images of the Lytro family of cameras. Finally, the key parameters of the MLA are tabulated that will be used as basic information regarding the MLA during the calibration process defined in chapter 7.

## 4 Parameters of Microlens Array

### 4.1 Introduction

The conventional cameras are precisely designed devices that capture the 3D object on a 2D photosensor in a single operation with various features that can be tailored to a specific application. The images recorded on the photosensor describe the captured scene in terms of height and width of objects, i.e. by knowing pixel resolution and number of pixels within object boundary, height and width of the object are expressed in SI units. But the 3<sup>rd</sup>-dimensional data (depth) is lost, however, indirectly the depth data can be generated by measuring the changes in object height or width. To retain the 3<sup>rd</sup>-dimensional data with the help of the 2D photosensor requires extra effort than just capturing a scene with a single exposure. Many research papers have been published suggesting possible ways to record the depth information such as; changing focal length, or changing object distance from the camera, and taking multiple images of the same scene [6, 54, 64, 65]. This procedure helps to detect changes in object size in every image and hence depth details can be predicted.

LF imaging systems are based on a conventional camera model to record 2D information along with additional details that help to generate depth details [13, 66]. In LF system, the 3<sup>rd</sup>-dimensional information is recorded in a passive way as a function of MLA encoded image plane data instead of using to record depth details directly, LF cameras record the direction of light rays that can be used to generate required depth data. For such purposes, a group of additional optical elements are used in LF cameras that facilitates the required information needed to generate depth data from a 2D photosensor, known as the microlens. The microlens is a very small optical element typically made from glass (other materials are not considered) with a property of bending the incoming light rays towards the optical axis. This property is identical to convex lenses [71] used in many instruments and cameras. As previously described, the microlens has a well-defined focal length ( $f$ ) and surface area, and when microlenses are combined results in single MLA. The MLA divides the bundle of incoming light rays into smaller sections, bends each section of light rays towards the optical axis of individual micro lenses and generates multiple bundles of light rays with minimal loss in information [68, 69]. Before the introduction of the MLA in the field of LF

photography, these optical elements were extensively used in CCD arrays [70, 71], digital projectors [72–74] and photovoltaics [75, 76].

## 4.2 Raw image of light field camera

The MLA is an important optical element that enables the 2D photosensor to capture 4D LF data in a single operation. The interaction of light rays with a conventional camera and an LF camera is the same with respect to the primary lens, where the primary lens performs the function of bending light rays towards the optical centre. However, when an MLA is placed in the path of light rays towards the photosensor, additional information of direction is recorded on the photosensor. The microlenses of the MLA are very small in size compared to the primary lens (Lytro-I generation camera's primary lens is approximately 140 times larger than the MLA size, the MLA is approximately 4.5 mm x 5.4 mm), hence the light rays passing through the MLA fill a small area of pixels in the photosensor and a group of such pixels is known as a microimage.

Due to the slight modification in the design of an LF camera with respect to conventional cameras, the raw images from these cameras are entirely different. This effect is shown in Figure 4.1, where the raw image of the regular camera resembles the captured scene without any further processing or alterations of pixels by the user. Whereas the raw LF image does not resemble the scene because it is a collection of light rays from different directions, bundled as a group classified based on their native microlens. To cause a raw LF image to resemble a real scene, further software-based pixel adjustment (eg. Spatial adjustment) is necessary (generating perspective images, see Chapter 6). The raw image from the conventional 2D camera is restricted to a single view, i.e. the camera records the scene normal to the optical axis of the camera, whereas the LF camera records the raw image as a collection of images with different views and the number of views depends on the microimage of the MLA. Figure 4.1 shows the Lytro-I generation camera images, that has a 10 x 10 microimage and hence 10 x 10 different views can be seen in each hexagonal grid of the raw image.

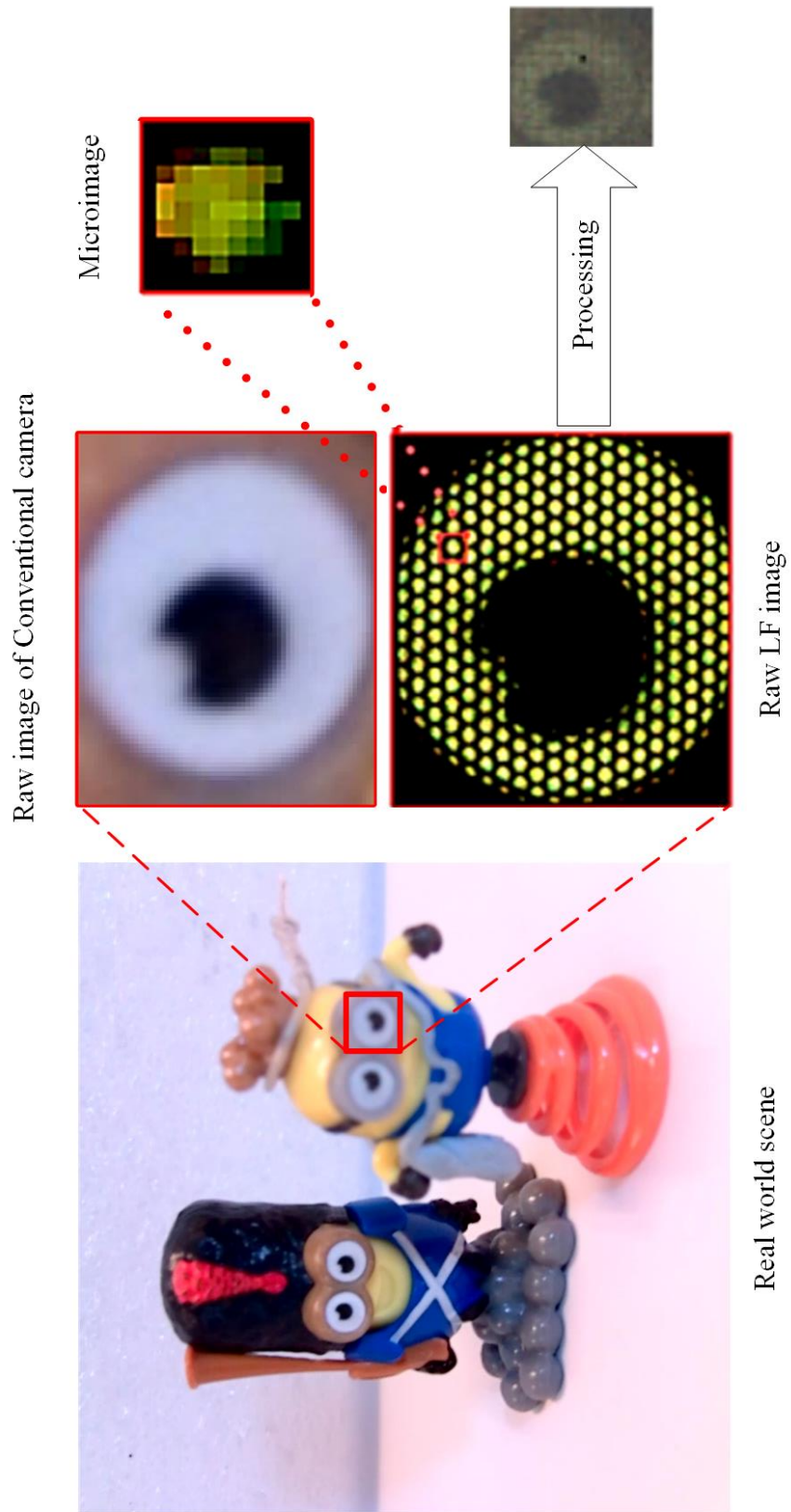


Figure 4.1 Light field camera image versus conventional camera image



In the case of normal 2D cameras, the final pixel count of the image depends on the number of pixels in the photosensor, i.e. the final pixel count of the conventional image is close to the pixel count of the photosensor used to record the image. In contrast, in LF camera, 4D data is recorded on a 2D sensor and hence the final pixel count of the image is less than the total pixel count of the photosensor. For example, the Lytro-I generation cameras use approximately 10 megapixels (MP) sensor to record LF data but the final raw 2D image available is of 0.1 MP as central camera view (Lytro software enhances 0.1 MP single 2D view Lytro data to 1 MP by image interpolation between neighbouring sub-aperture images). By sacrificing spatial pixel count of the final image, LF cameras record different views of a scene over a single exposure [16, 17]. Hence the pixel count of an LF camera image (central view) is given by Equation 4.1.

$$Pixel\ Count = \frac{Sensor\ pixel\ count}{Number\ of\ pixels\ in\ a\ Microimage} \quad 4.1$$

For example, the Lytro-I generation camera generates a raw data file of 3,280 x 3,280 pixels (10 MP), while every individual microlenses use 10 x 10 pixels to record data (known as microimage size) and hence the final image resolution of a single view is 0.1 MP. A limited number of research results have been published defining methods to increase the final resolution of LF cameras. A new LF camera model was proposed for this purpose, called Plenoptic 2.0, which increases the final resolution up to 9 times the original resolution of the LF camera used [63]. But this new method generates a 2D image as a conventional camera by interpolating positional and angular information available in captured radiance data, thereby increasing the final pixel count of the final image. In this method, the MLA was focused on the image plane formed by the main lens, rather than on the main lens (see chapter 2, section 3.2). For generating depth maps from an LF camera, both direction and intensity information (4D data) is essential and hence in this work the aim is towards generating depth map in absolute scale at first instance. Therefore, more preference is given overall 4D data over the resolution of output 2D RGB data since resolution along with the quality of images can be increased by interpolating multiple microlens data.

Very few published articles explain the significant effect of focal length, size of MLA on image quality, and overall performance of the LF camera. However, general understanding can be drawn with some known properties of conventional cameras and their optical elements. The number of light rays recorded in an LF image depends upon the ratio between active and passive regions of the MLA. The area of the MLA that serves as the lens and performs the light bending task is known as an active region, while the regions in between two or more active regions are known to be passive regions (do not take part in bending of light rays). In Figure 4.2, two varieties of MLA's are shown which are typically used in LF cameras. The hexagonal Type-2 configuration has a high ratio of active region over passive region when compared to Type-1. While for spherical MLA configurations both the types have the same amount of passive region due to the minimum package density of the convex spherical lenslet.

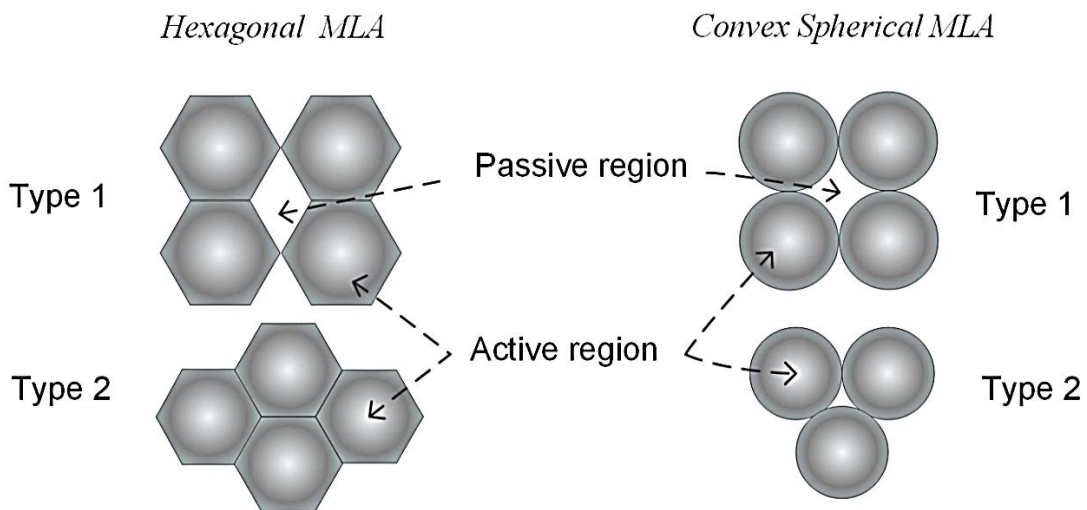


Figure 4.2 Different regions in a Hexagonal and Spherical MLA

The focal length of the MLA also plays an important role in recording LF data onto a photosensor since an aberration-free MLA helps to bend all incoming light rays towards the optical axis of individual microlenslets. The field-of-view (*FOV*) and the angular resolution are among the important terms to be considered while defining (selecting) MLA's focal length for an application. The MLA with a shorter focal ( $f_1$ ) length will record larger data scenes (light rays) using photosensor space ( $h$ ) (see Fig 4.3) and hence higher angular resolution for image processing. For example, Lytro I-generation cameras used an MLA of 25  $\mu\text{m}$  focal length that is better compared to a Thorlab 1.24 mm focal length MLA for gathering higher angular data. Also, by selecting a

photosensor of small pixel size, the number of pixels in each microimage can be increased. For example, consider a photosensor with  $1.64 \mu\text{m}$  pixel size is replaced by a photosensor with  $2 \mu\text{m}$  pixel size, thereby providing more area (pixels) to record angular data as shown in Figure 4.3. The MLA with a longer focal ( $f_2$ ) length tends to have less angular information because the FOV is minimised for the same photosensor space ( $h$ ). In this situation, post-focusing capability decreases, since light rays tend to become more parallel to each other as they travel a long distance from the MLA.

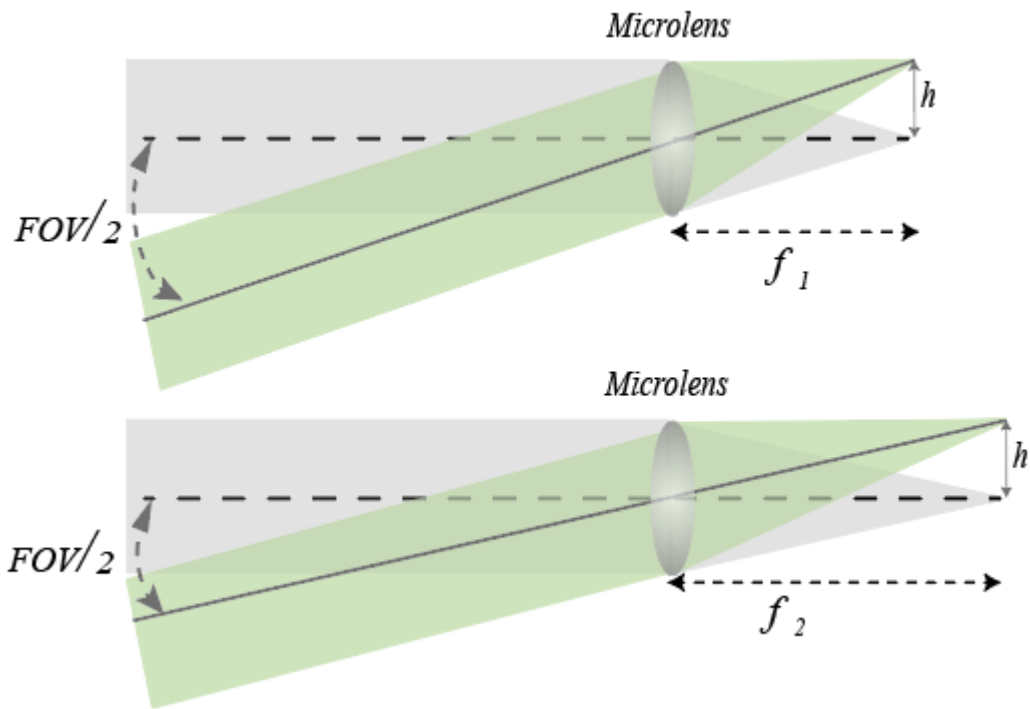


Figure 4.3 FOV, angular resolution and focal length of a microlens array

All the light rays passing through a lens can be re-traced using ray diagrams by knowing details such as the object distance from the lens, the imaging plane distance, and focal length of the lens (variables of the thin lens equation [19, 20]). Similarly, by knowing the focal length and active region of the individual microlens, the interaction of light rays with the MLA can be retraced. To achieve this goal, all the pixels in every microimage must be addressed by their own microlens, i.e. the microlens number is used as an address to find the relevant pixel in a microimage.

The active region of the microlens is expressed in terms of pixel count given by Equation 4.2, which is the total number of pixels under each microlens. Let  $p_x$  and  $p_y$  represent the pitch of the microlenses in  $X$  and  $Y$  direction respectively as shown in

Figure 4.4, while  $\bar{f}$  represents the focal length of the MLA as shown in Figure 4.5. These factors, in turn, decide the quality of the image and amount of light recorded by each group of the microlens depending upon the distance between the MLA and the photosensor. By knowing  $\bar{f}$ ,  $p_x$ ,  $p_y$  and the distance of the MLA to the photosensor, the angle of incoming light rays can be calculated.

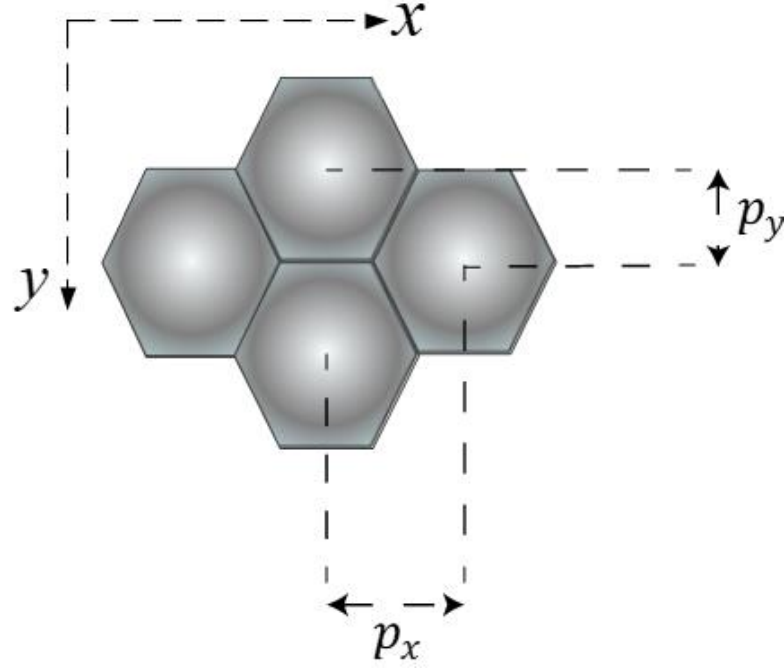


Figure 4.4 Pitch of microlens array

Let  $(u_i, v_i)$  be the microlens considered with absolute distance  $d''$  from photosensor and,  $I_x$  and  $I_y$  be the pixel under consideration that belongs to the microlens  $(u_i, v_i)$  as shown in Figure 4.5. Under the given situation of  $d'' = \bar{f}$ , the angle of incoming light rays  $\theta_{xy}$  are directly proportional to the distance of the pixel from microlens centre  $(I_{cx}, I_{cy})$  given by Equation 4.3.

$$\text{Pixels in microimage} = \text{MLA pitch/pixel size} \quad 4.2$$

$$\theta_{xy} = \tan^{-1} \left( \frac{(I_{cx}, I_{cy}) - |(I_y, I_x)|}{f} \right) \quad \forall (I_y, I_x) \in (u_i, v_i) \quad 4.3$$

To accomplish the task of finding the angle of the incoming light rays through the MLA and access other LF features, it is important to know the location of the principal ray of each microlens in terms of the pixel (see Figure 4.5). Since the MLA is placed on the photosensor as an additional unit, there is no direct relationship between the MLA principal axis and photosensor pixels. So, image processing techniques are used to calculate the principal axis of MLA elements (MLA cord) (see Sec 3.4.2).

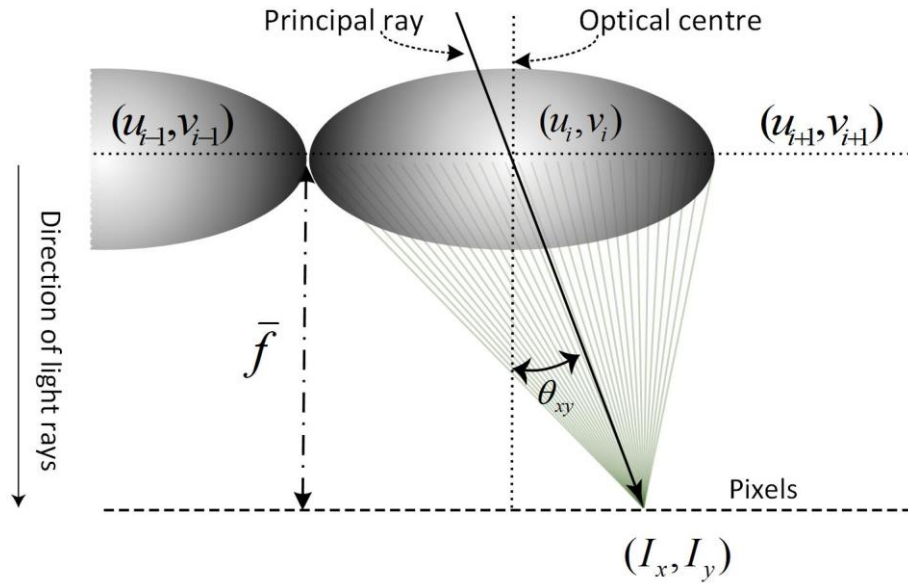


Figure 4.5 Angle of incoming light rays

### 4.3 Geometric and light ray dependent MLA cord

The MLA cord is one of the key pieces of information required to access the main features of LF data. The cord data helps to identify the microlens principal axis which is a key element to transform the microlens into a pinhole model to trace the light rays passing through it. It also plays a key role in the splitting of LF data into 2D camera views (see Chapter 5) for calibration and depth generation. Also, the principal ray directly depends on the direction of incoming light rays.

The optical centre of a lens lies on the plane where the incoming light rays, perpendicular to the lens plane meet up. The light ray passing through is called the principal ray and the distance to the photosensor is approximately equal to the focal length (in case of LF cameras). To analyze the situation of light rays from the MLA converging on the photosensor, microlenses can be represented as pinhole models where a single light ray (beam) can pass through any given pinhole array. These pinhole

arrays generate corresponding pinhole centres on the photosensor which depends on the microlens distance from the optical centre of the main lens. These pinhole centres are also known as MLA cords and finding the pinhole centre as described above in the context of this thesis is known as the light directed method. It should be noted that the pinhole centres depend on the location of principal ray and not the physical structure or shape of the microlens.

Some of the existing research tends to calculate the microlens principal axis to sub-pixel accuracy by considering geometric hexagonal patterns [81, 82]. With the geometrical structure method, the optical centre of the microlens is considered as the principal ray irrespective of microlens location from the optical centre of the camera. Whereas the principal ray shifts away from the optical centre along with the distance from the optical centre due to distortion of lenses [86]. The MLA cord found using geometrical structure and light directed methods differ in the order of 0 to 2 pixels with respect to the Lytro family of cameras. Due to the principal axis shift, there is a significant improvement while generating LF features using light-directed MLA cord method over geometrical structure method. The shift of 0 to 2 pixels is significant for some metrological applications where LF camera is used for measuring distance. Since every shift in the pixel location tends to change the final depth results. If being used for generating scene capture (i.e. photograph) relative depth map shift of 0-2 pixels is generally not too important since RGB data is generated by interpolation with neighbouring pixels of sub-aperture images.

Figure 4.6 compares the two scenarios of geometric versus light directed MLA cord models, with the shift in the principal axis of the microlens as the distance increases from the principal axis of the main lens ( $o$ ). This shift in principal axis is (as previously stated) approximately 2 to 3 pixels and may increase under a serious distortion of the main lens. The light rays considered for the demonstration are parallel to the optical axis of the system (main lens + MLA). Since the MLA is assumed to perform as a pinhole array, co-linear light rays would be a good choice to identify the principal axis of MLA. When the geometric cord system is selected, it can be observed from Figure 4.6 that the optical centre of microlens ( $o'$ ) remains invariable with distance from the main lens principal axis  $o$ . This situation considers the light rays passing through the

MLA are co-linear to individual microlenses and hence fails to generate the principal axis shift.

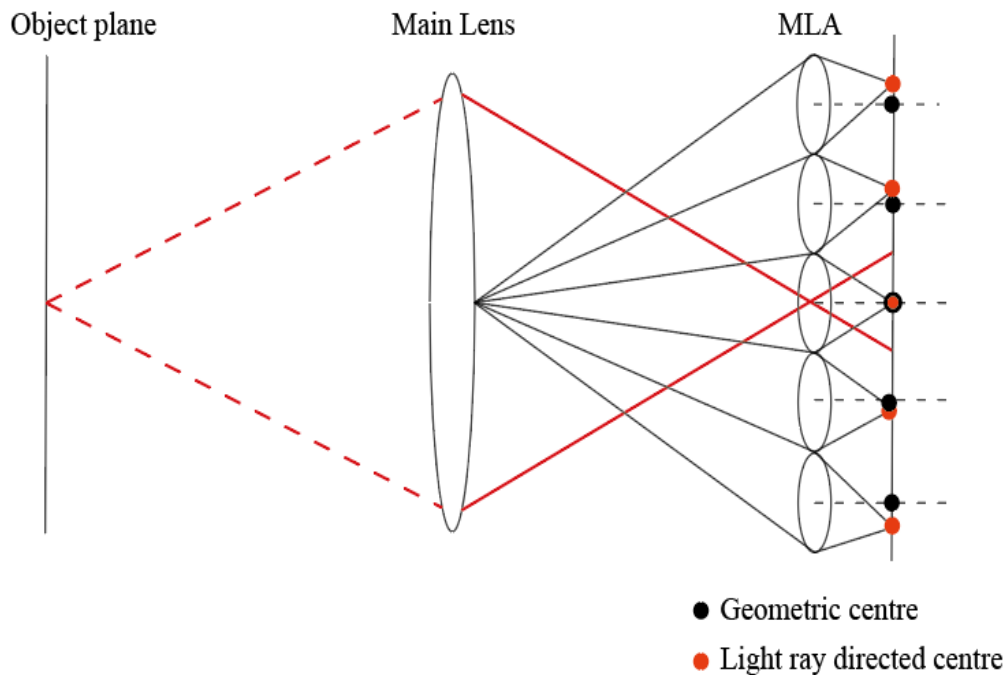


Figure 4.6 Geometric vs light directed MLA cord model

The pixel corresponding to the optical centre of the MLA is considered for generating the MLA cord represented by the black dot in Figure 4.6 (this follows the geometric centre method). In contrast, the principal ray shifts a few pixels away from the microlens geometric centre represented by the red dots in Figure 4.6. Understanding this shift while selecting microimage pixels, helps to recover uniform data throughout the MLA plane. In this situation, the rate at which principal ray shifts from the MLA optical axis is directly proportional to the radius of the main lens and  $\bar{f}$ . The MLA cord generated using the light directed method ensures the required shift of the principal ray that generates accurate sub-aperture images during the process of depth calculation.

The difference between geometrical centre method and light directed method is represented using an entire raw image that was used to calculate MLA cord for the Illum camera is shown in Figure 4.7 and is calculated using the procedure illustrated in the Sec 4.5. At every microlens centre, the difference between the two methods is highlighted as that range between 0 to 2 pixels for Lytro Illum camera. It can be noticed that in the error map that the maximum principal shift is 2 pixels and the error gradually increase along with the distance from the centre of the image (since the

decimal values are converted to integer values, the Figure 4.8 looks pixelated and non-uniform). A similar error map for the Lytro-I shows worse results and hence not shown.

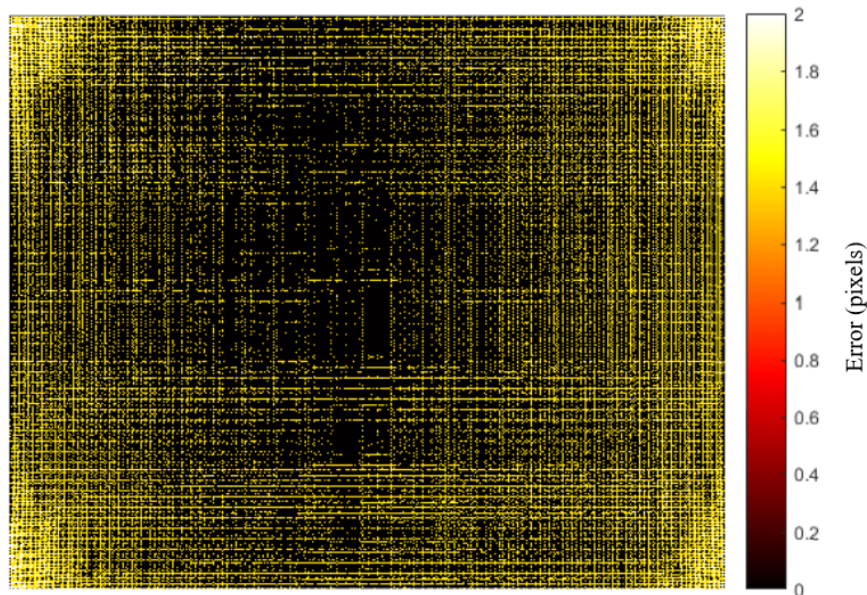


Figure 4.7 Shift map(pixels) of Geometrical cord vs light directed cord method for Illum camera

#### 4.4 Mounting issue of MLA on Photosensor

Microlens arrays are placed in between the primary lens and the photosensor to capture LF information in a single operation. The focal length of the MLA is very small (25  $\mu\text{m}$ ) compared to the focal length of the main lens (6 mm) and hence the MLA is mounted directly on top of photosensor with the help of spacing elements. The thickness of these spacing elements is very close to the focal length of the MLA. The spacing elements can vary in shape, size, and materials. The Lytro I-generation camera is provided with polymer-based spacing elements. Care must be taken while mounting the MLA on the photosensor because unwanted misalignments cause problems with recording LF data at the highest quality.

There are two types of misalignment that can be introduced while mounting the MLA on the photosensor; tilt, and rotation. The distance between the MLA and the photosensor plays a key role in recording LF data, but due to tilt introduced while mounting the MLA can cause a major problem which results in variable distances between MLA and photosensor as shown in Figure 4.8.



A slight offset in the  $(u, o)$  plane with respect to the photosensor plane  $x$  results in a tilt of  $\beta$ . While offset in the  $(v, o)$  plane with respect to the photosensor plane  $y$  results in a tilt of  $\alpha$ . It can be noticed from Figure 4.8 that misalignments are split up into  $x, y$  planes, but in the practical scenario, the tilt misalignment is solved along with lens distortion during the calibration process and is explained in Chapter 7.

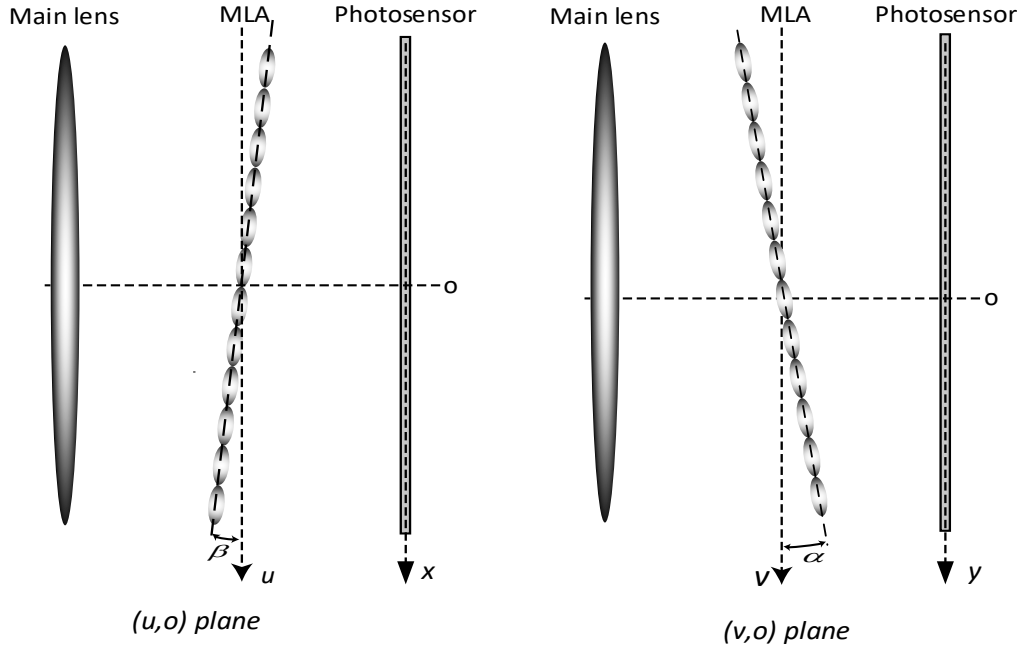


Figure 4.8 Misalignment in X and Y plane – Tilt

The rotational misalignment is the second error introduced in the process of mounting the MLA and this alignment issue does not affect the final data recorded by the LF camera. But addressing this error before generating the MLA cord will ease MLA cord detection and allotting pixels for accessing LF features. Figure 4.9 (top) represents a perfect photosensor-MLA assembly without any alignment issues. Whereas, Figure 4.9 (bottom) shows the rotational error in  $(u, v)$  plane resulting in  $\theta_x$  and  $\theta_y$ , that indicate the error introduced with respect to the photosensor plane  $(x, y)$ . The error  $\theta_x$  and  $\theta_y$  can be positive or negative depending upon the mounting issues and is calculated using Equation 4.4 where  $r$  and  $c$  represent the row and column pixels while  $\Delta r$  and  $\Delta c$  represent the changes in  $r$  and  $c$  respectively. .

$$(\theta) = \tan^{-1} \left( \frac{\Delta r}{\Delta c} \right) \quad 4.4$$

In this section, the shift error that is caused when the origin of MLA  $(0,0)$  is not matched with the origin  $(0,0)$  of the photosensor is neglected. In the presence of shift

error, some of the microlenses lie outside the active region of the photosensor and do not contribute to the LF acquisition. However, shift error does not affect the recording of LF data but introduces issues during processing of LF data.

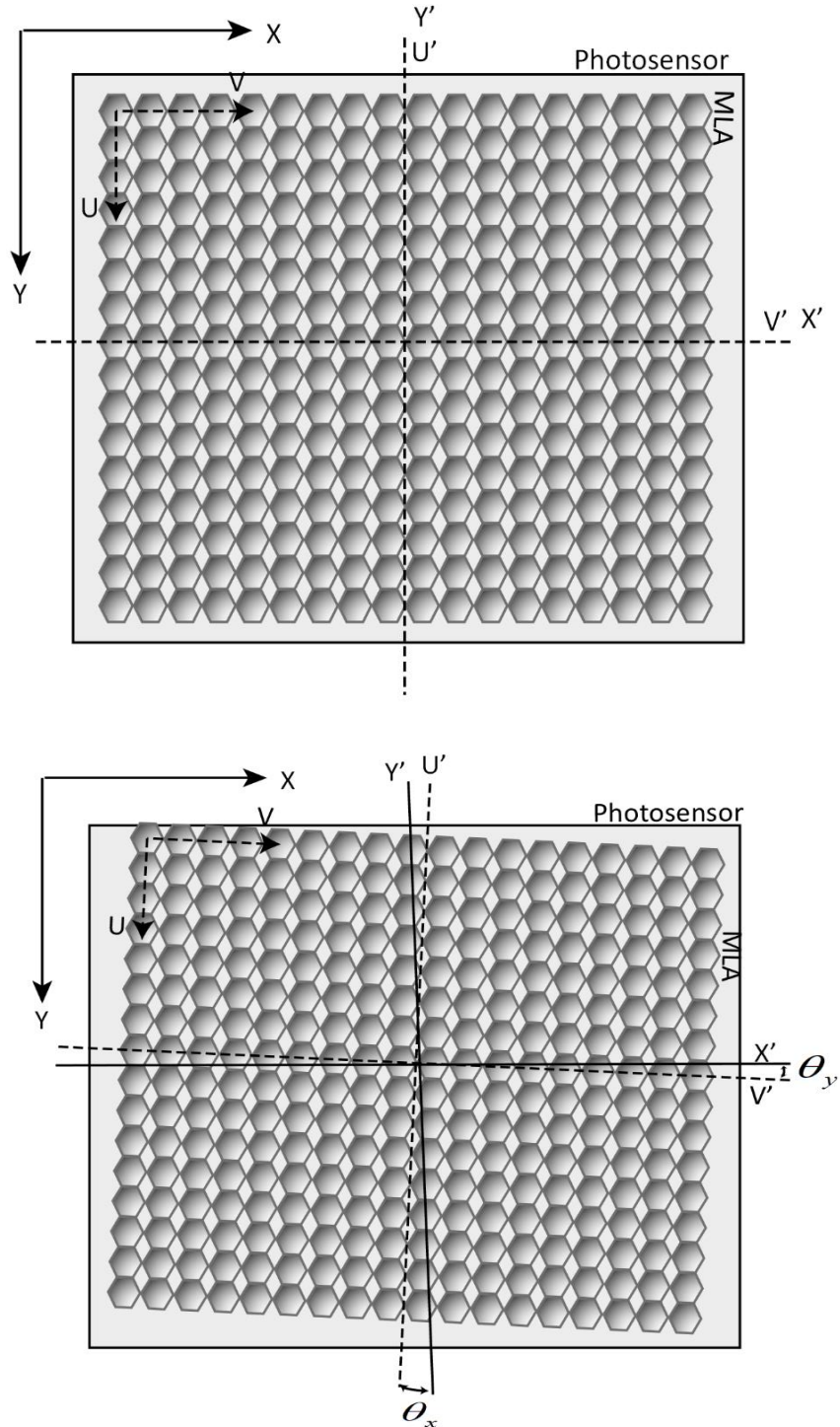


Figure 4.9 Perfect alignment of photosensor-MLA assembly (top) versus rotational misalignment  $\theta_x$  and  $\theta_y$  (bottom)

## 4.5 Algorithm to generate the MLA cord

As discussed in section 4.4, the MLA cord is essential for accessing many features of the LF cameras. The MLA cord represents the microlens principal axis in terms of photosensor pixel units, thereby generating a mathematical relationship between the MLA and the photosensor.

To generate the MLA cord using raw images of the LF camera is sometimes difficult because similar information is stored in each microimage pixels, as it represents a perspective view of the same real-world scene. To ease the MLA cord detection, it is important that radiance on the main lens is uniform.

---

### Algorithm 1: Automatic MLA Cord

---

```
1: procedure  $W = \sum_{i=1}^n w_i/n$ 
2: Convolve W with a 3x3 window averaging filter =  $\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ 
3: for (threshold=0:0.05:1) do
4:  $\arg_{max} \|count(segements)\|$ 
5: end
6: for ( $\theta=-1:0.05:1$ ) do
7:  $seg_{val} = \min(bw\_region)$ 
8: end for
9: find  $p_x, p_y, I_{cx}, I_{cy}$ 
10: Apply boundary constraints
11:  $int_{cord} = seg_{val}(\theta)$ 
12: return cord parameters
```

---

Figure 4.10 Algorithm to find Automatic MLA cord using the light-directed method

Hence from a practical viewpoint, the main lens is covered using a semi-transparent (diffuser) or the camera is placed normal to an LCD monitor emitting uniform white

light. The raw image captured using this arrangement is called the white image ( $w$ ), which is used to determine the MLA cord of the LF camera with the help of image processing algorithms (such as the one documented in Figure 4.10) and techniques.

The elements of the algorithm are discussed as follows. Let  $n$  be the number of white images ( $w$ ) used and the final averaged white image  $W$  is given by  $W = \sum_i w_i/n$  (line 1, see Figure 4.10). The image  $W$  is smoothed using an averaging filter of 3x3 window size and later normalised between the values of 0 and 1 (line 2, see Figure 4.10). For efficient automated segmentation of the microimages, a threshold value needs to be calculated that classify  $W$  image into a useful binary image (line 3 to 5, see Figure 4.10). In the binary image, pixels with a value of 1 represent the microimages and pixels with a value of 0 represent the boundary of each microimages as shown in Figure 4.11.

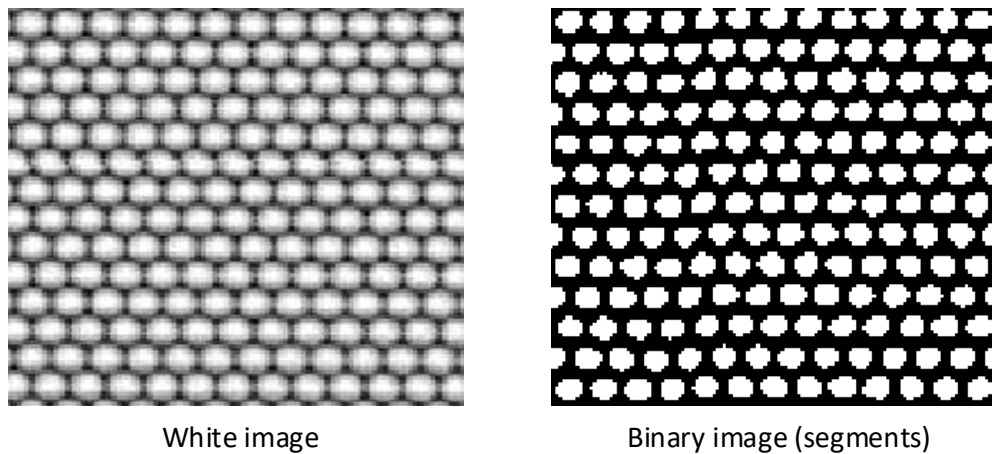


Figure 4.11 Representation of white image and corresponding binary image for a given threshold value

Using Equation 4.5 over the binary image for different threshold values generates an optimal value resulting in a maximum number of segments (microlens). The total number of segments found is not always equal to the number of microlenses present in the MLA. This is due to misalignment issues as explained in the sec.3.5. The tilt in the MLA causes variable amounts of intensity representing individual microlens over the entire photosensor plane and hence fails to show the total segment value. Also, rotation error forces light from microlens away from photosensor active surface and hence total segmentation count is lower than the actual microlens count in the MLA. The threshold values for the Lytro family of cameras is shown in Figure 4.12 and Figure 4.13. with

threshold values of 0.55 and 0.60 for Lytro-I and Illum respectively (line 6 to 8, see Figure 4.10).

The total number of the microlens in the Lytro-I and Illum cameras are approximately equal to 107,584 (328 X 328) and 236,958 (434 x 541), while the total number of cord found using Equation 4.5 at the corresponding threshold values are 124,177 and 234,794 respectively.

From the threshold graph for the Lytro family of cameras, it can be noticed that Lytro-I has gradually increasing segment count compared to the Illum camera, indicating the quality of light spread over the photosensor plane. In the Lytro-I camera, vignetting causes intensity loss which is directly proportional to the distance from the optical centre of the main lens. The Illum camera is designed with a highly efficient and uniform light spread lens system compared to the Lytro-I, and hence higher price margin for Illum cameras.

$$threshold = \arg_{max} |count(segements)| \tag{4.5}$$

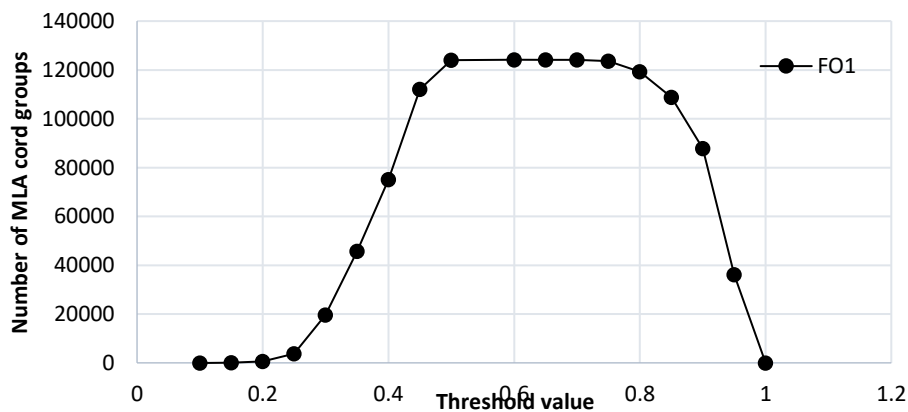


Figure 4.12 Threshold value vs Number of cords detected- Lytro-I

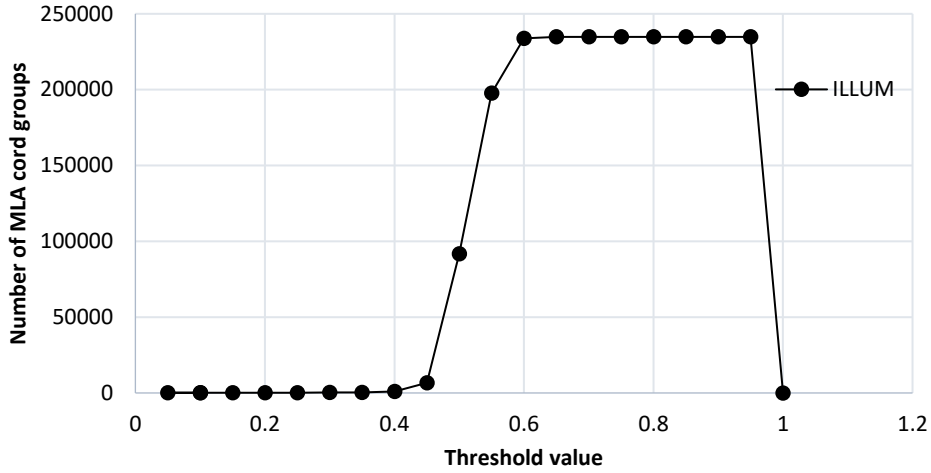


Figure 4.13 Threshold value vs Number of cords detected-Illum

Further image processing techniques are applied to the binary segmented image generated by using the calculated threshold value. The next step is to sort out the rotational error where the MLA plane is made parallel to the photosensor plane. The rotational error value (in degrees) is calculated by randomly rotating the image under consideration to obtain a minimum number of segments in each column. The binary segments are detected and a pixel with maximum intensity will represent the segment location. In the absence of rotational error, the segments (microlens) in the  $v$  plane is parallel to the  $y$  plane (see Figure 4.9). Counting the number of segments in every column of the photosensor with respect to the maximum intensity pixel and averaging by the number of columns in the photosensor, will result in the actual number of microlenses in every individual column of the photosensor under 0 degrees tilt. The binary image is rotated by a known value of  $-1:0.5:1$  degrees and the resulting ratio of segment count and number of columns is calculated to reach a minimum value ( $rotation_y$ ) using Equation 4.6. This procedure is also applied in the  $x$  direction to generate a  $rotation_x$  value using Equation 4.7.

In Figure 4.15, the rotational error values plotted against a number of microlens segments found in both the  $x$  and  $y$  directions are shown. It can be observed that approximately 3,300 segments were found when rotational values are in the range of  $-1.0$  to  $0.25$  degrees and  $0.7$  to  $+1.0$  degrees. Since the Lytro-I has in built rotational error (manufacturing and assembly, as observed by [87]), introducing additional rotational value caused the segment count to be constant at 3,200 indicating the presence of at least one microlens segment in every column of Lytro-I photosensor of size  $3,280 \times 3,280$ . However, when the rotational value introduced is between  $0.2$

degrees and 0.65 degrees, the segment count drops very close to 1,700 and 856 in the  $x$  and  $y$  directions respectively, indicating that the introduced rotational value nullifies existing rotational error of Lytro-I. The MLA count of Lytro-I is 376 x 328, while segments detected are in the range of 1,700 and 856. The higher values of segment location demonstrate that the microlens principal axis is shifted and hence adding up a number of column count, i.e. the MLA cord location of microlenses are shifted and do not fall in the same line, hence boosting the cord count number. So, a rotational error of  $0.35^\circ$  can be noticed in the Figure 4.15.

Similar results for the Illum camera are shown in Figure 4.16, where the microlens segment count starts approximately at 5,200 and 5,400 in the  $x$  and  $y$  directions. Since there are no mounting issues with the Illum design, when the rotational error is decreased from  $\pm 1$  to 0 degrees, the segment count value drops down to 2,900. The approximate MLA count of Illum camera is 434 x 541, while the segment count is in the order of 2,900 indicating a shift in MLA principal axis in terms of pixels to boost up segment count see Figure 4.14).

$$rotation_x = \arg_{min} \left\| \sum_{c=1}^n count(segments)_{r=1:m} \right\| \quad 4.6$$

$$rotation_y = \arg_{min} \left\| \sum_{r=1}^m count(segments)_{c=1:n} \right\| \quad 4.7$$

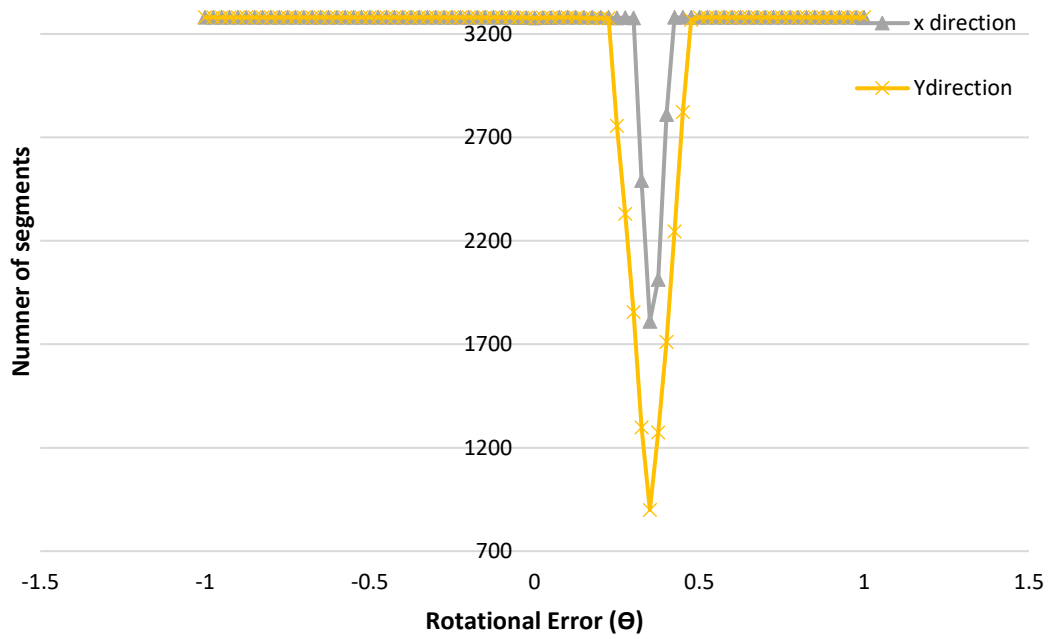


Figure 4.15 Angle vs Cord detected – Lytro-I

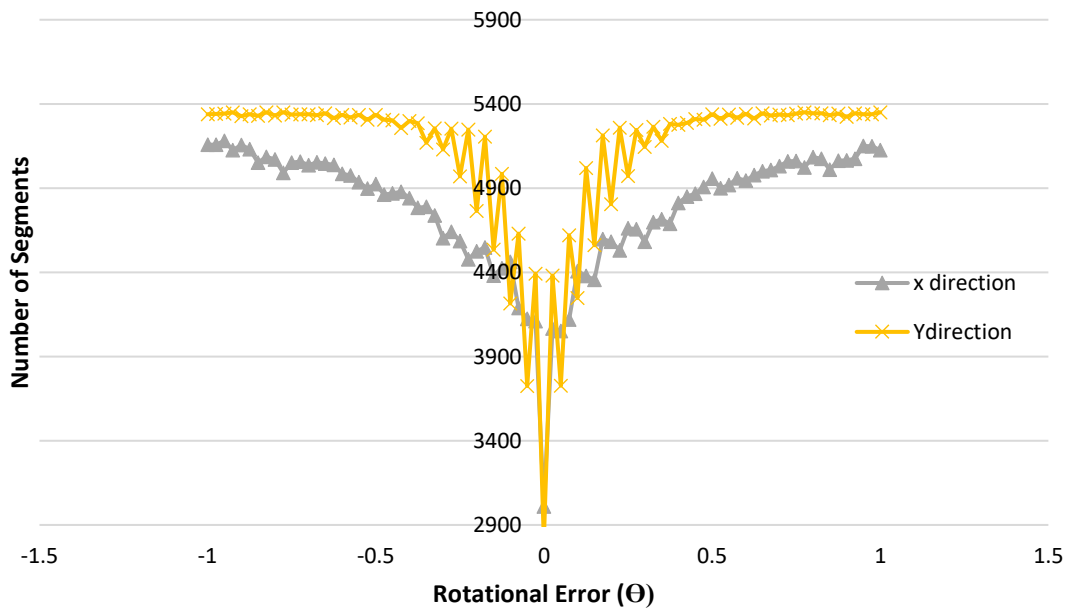


Figure 4.16 Angle vs Cord detected – Illum



In the process of determining the rotational error, the average pixel distance between all the segments, the pitch of MLA ( $p_x, p_y$ ) in the  $x$  and  $y$  directions is calculated as defined in Equation 4.8 (line 9, see Figure 4.10), where  $r$  and  $c$  represent the row and column values. Also, the MLA pitch acts as a boundary condition to make sure that all cords detected have a minimum number of neighbours in terms of pixels, which is equal to (pitch x pitch), and has valid pixel locations (Equation 4.9) (line 10, see Figure 4.10). In Equation 4.9,  $I_{cx}$  and  $I_{cy}$  represent the cord centers in  $x$  and  $y$  directions respectively,  $I_c$  represents the overall MLA cord indicated by ( $I_{cx}, I_{cy}$ ) format.

$$p_x = \frac{(\sum_{i=1}^n \text{cord}(r, c) - \text{cord}(r, c + i))}{\text{length}(\text{cord})} \quad \forall i \leq m$$

4.8

$$p_y = \frac{(\sum_{i=1}^n \text{cord}(r, c) - \text{cord}(r + i, c))}{\text{length}(\text{cord})} \quad \forall i \leq n$$

$$I_{cx} = \begin{cases} \text{cord}(c) & \text{if } (r \geq r_i + p_y) \\ 0 & \text{otherwise} \end{cases}$$

4.9

$$I_{cy} = \begin{cases} \text{cord}(r) & \text{if } (r \geq r_i + p_y) \\ 0 & \text{otherwise} \end{cases}$$

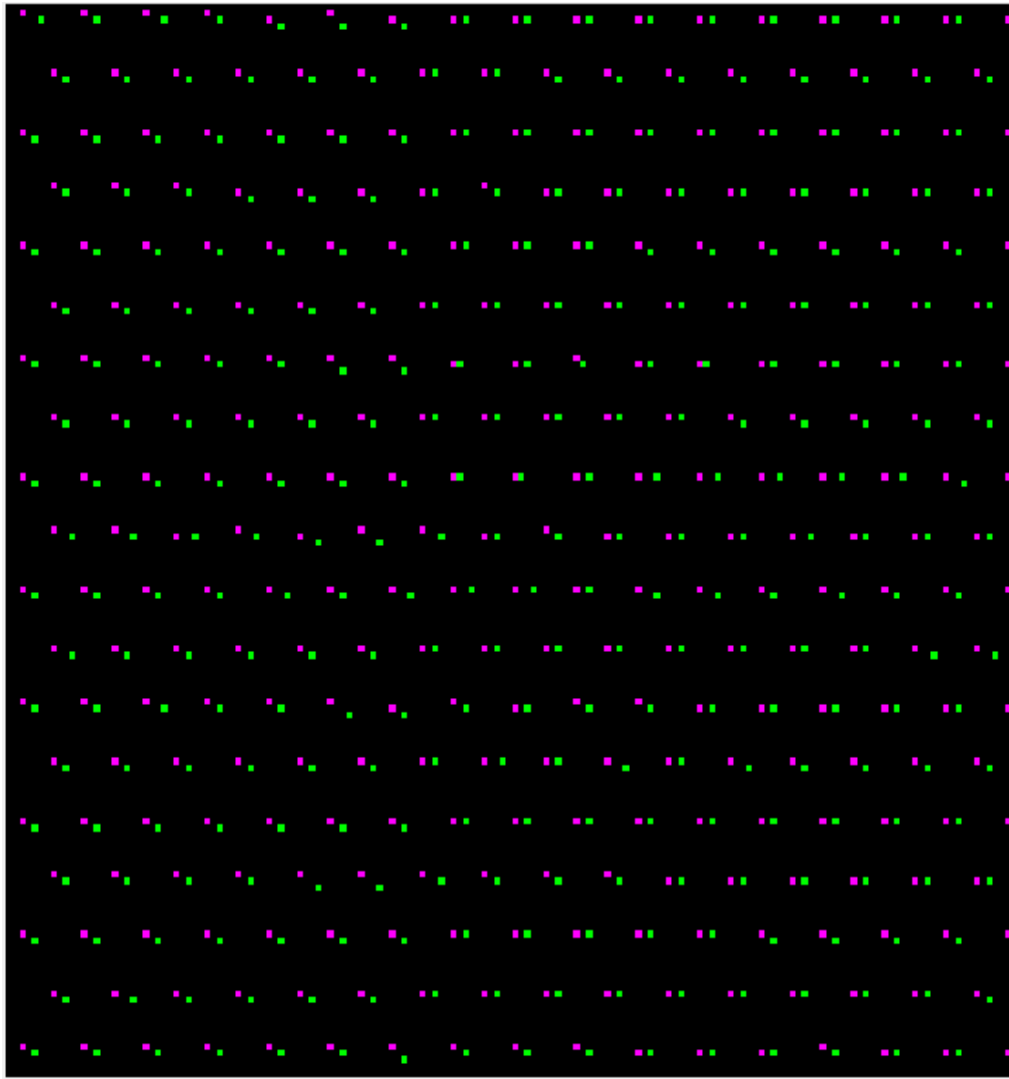


Figure 4.17 Representation of two cord detection results

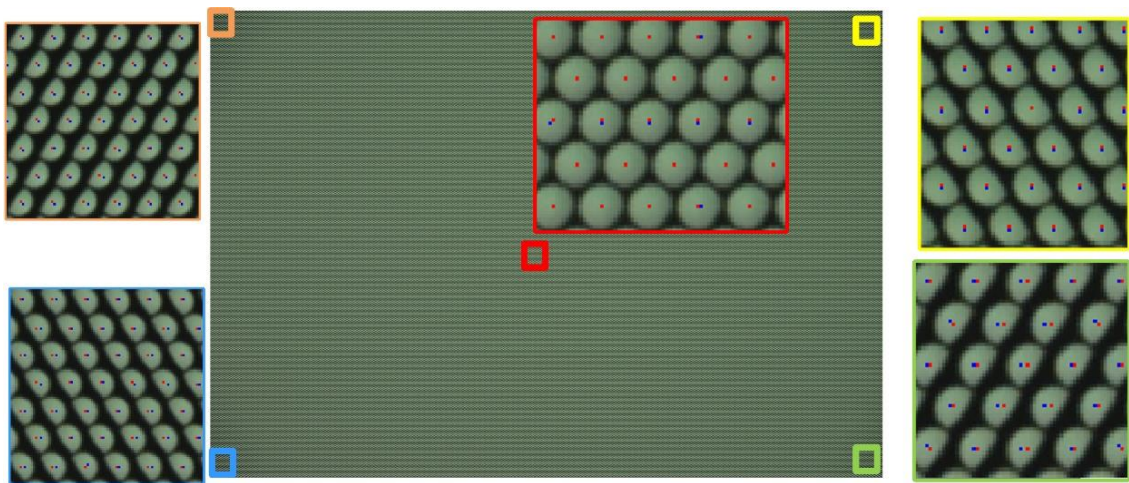


Figure 4.18 Results of light directed cord detection for the Illum cameras white image

Figure 4.17 represents the MLA cord detected using the light directed method (red dots) following the algorithm represented in Figure 4.10 compared with geometrical cord detection method (green dots) (see Sec 4.3). A small section of the entire MLA results are shown and the shift between both the method are clearly visible. It can also be observed in Figure 4.17 that red dots seem to be non-directional and shifts are not concurrent, i.e. all new cord detected (red dots) does not shift in one direction and quantity of shift vary. This is because, the tilt of the MLA is not considered in this chapter, as tilt will be corrected in the calibration (radial+skew) procedure (see section 4.4). The overall results for the Illum camera white image are shown in Figure 4.18, and the enlarged view highlights the clear view of MLA cord detected at the centre and edges of the photosensor.

## **4.6 Conclusion**

The presence of an MLA in the optical path of the camera, in between the main lens and the photosensor generates additional angular data along with intensity data. For accessing many features available with the LF cameras, some basic parameters of the MLA must be gathered by the user. Some of this information is provided by the MLA manufacturer, while the rest of the parameters must be calculated by the user. In the process of which many issues need to be solved by the user to calculate error-free MLA parameters. In this chapter, some of these issues have been explained theoretically along with some novel contributions.

The basic understanding of LF cameras illustrated in Chapter 3 has been narrowed, focusing on the function, property and parameters of the MLA in this chapter. The key role of the MLA in an LF camera has been discussed in this chapter along with other properties such as differences between conventional raw images and the LF image, the effect of focal length on LF data and direction of light rays in terms of pixels.

Furthermore, the major difference between the geometrical MLA cord and light directed MLA cord has been discussed in this chapter. The MLA cord detected with these two methods differ only by 2 to 3 pixels but has a major impact while accessing LF camera features for metrological applications where pixels represent distance values. The MLA cord detection algorithm described in this chapter identifies MLA pitch using the light directed method. The misalignments that are introduced while

mounting the MLA on the photosensor are addressed in this chapter, along with resolving rotational issues.

The MLA cord generation algorithm (Figure 4.10) discussed in this chapter is not limited by the shape of the MLA (square, hexagonal) compared to existing results [30, 85]. Moreover, the algorithm does not require any user input to generate the MLA cord, as required by existing results where the MLA pitch distance value has to be entered by the user [85]. Using this algorithm, basic MLA parameters such as  $p_x, p_y, I_{cx}$  and  $I_{cy}$  are evaluated demonstrating the first novelty of this work. Results of the algorithm used to calculate basic MLA parameters of Lytro family of cameras are given in Table 4-1.

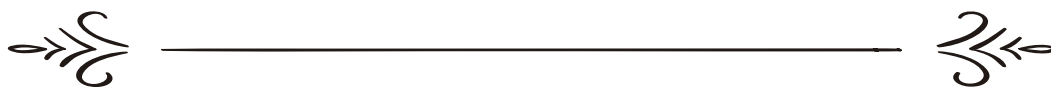
Table 4-1 Property calculated for Lytro family of cameras

Property	Lytro-I	Illum
MLA Pitch (Pixels)	10x10	13x13
MicroLens count	376x327	434x 541
Rotation error (degrees)	0.35	0

The information and understanding obtained from this chapter lead the way to further understanding about how LF data is processed, accessing features such as refocusing, splitting LF data into 2D camera views and later into generating depth maps. These features will be briefly discussed in the next chapters, 5 to 8.



# 5



ABSOLUTE DEPTH  
USING  
LYTRO CAMERA DEPTH MAPS

## *Overview*

The Lytro cameras were the first commercially available LF cameras that are low-cost and capable of capturing angular data along with the intensity, in a single operation. The Lytro company introduced two versions of LF cameras, specifically targeting the community of photographers to compete with standard DSLR cameras. However, due to the additional angular data captured by the Lytro cameras compared to the conventional camera along with low-cost margin, the Lytro cameras, are also of interest to the engineering community.

The computer vision community has previously used the results of the Lytro cameras, especially the greyscale depth map generated by the Lytro desktop software and added additional enhancements, to make the depth map look more realistic, although accepting the fact that the depth map is relatively scaled (not measured in SI units). Research in the field of metrology using low-cost LF cameras is limited, with no research specifically targeting the Lytro cameras. Thus, in this work a method to generate absolute depth data from the Lytro family of cameras is developed along with identifying key features such as; accuracy, repeatability and work volume being addressed, thereby defining the novelty of this work. Furthermore, response curves are introduced which define the relationship between depth (mm) to the Lytro depth (greyscale), and hence the sensitivity of the camera to external conditions.

## **5 Absolute Depth Using Lytro Camera Depth Map**

### **5.1. Introduction**

The Lytro company, USA, launched the first commercially available LF camera capable of recording both intensity and angular data with a single exposure. The cameras were specifically aimed at the consumer market with two variants; Lytro-I generation and Illum cameras, launching in the year 2012 and 2014 respectively [89]. These Lytro cameras are designed to target the consumer photography market to compete with conventional DSLR cameras giving great flexibility to the user with respect to focal length, aperture, exposure, and many other features required by a standard photographer.

The Lytro-I generation camera has great flexibility with automatic and user-selected values for exposure and shutter speed features, along with a touch to focus feature (TTF). This version of the camera is at the lower end of the cost margin for the LF cameras available in the market (costing around £100). Even though the Lytro-I generation camera was inexpensive, it was not recognised by the photography community because Lytro-I generation cameras lack some of the basic features such as variable aperture, storage capacity and familiar camera design features. The second version of the Lytro camera, the Illum, gave the user more flexibility and freedom over many camera features; such as the focal length, exposure and shutter speed as well as looking like a DSLR camera. Furthermore, the Illum was designed to offer many properties of the camera to the user on the go (before or while taking a photo) and the option of highlighting the depth information is one of the frequently used options among all including autofocus, focus to infinity and distance bar. These features encouraged the photography community to use the Lytro cameras and potentially replacing standard DSLR cameras. The Lytro company also offer a software tool that is used to view and access the LF images captured (Lytro Desktop Software (LDS)). The LDS can be used along with many existing software tools; such as Adobe Photoshop and other photo editing software, to enhance the image quality and details of the scene captured.



Apart from the photography community, the computer vision and machine vision communities are equally interested in the LF technology available in the commercial market. The Raytrix company is one of the first LF cameras manufacturers targeting engineering application users, providing different versions of their camera for specific applications. These cameras are expensive and on average each system costs £50k to £100k including the software package.

The underpinning technology used in the Lytro and the Raytrix cameras remains the same, with some minor changes with respect to the optical elements[12, 13, 87]. But the Raytrix cameras claim to generate depth maps in absolute scale while the Lytro cameras are designed to provide relative depth maps [88–90]. It must be noted that both the cameras rely on their own software tools provided by the respective companies to generate depth maps. To generate an absolute depth map using a Lytro camera, it is important to understand the properties of the camera along with the software tool (LDS). The Lytro company has not disclosed the algorithms used in the LDS to add user required enhancements (algorithms) that improve the output. The LDS is considered as a black box since the user has no control over the output and cannot manipulate the algorithms.

## **5.2. Features of the Lytro cameras**

The Lytro company provide the LDS to process LF images captured by their family of cameras. The LDS exports depth maps, all-in-focus images, perspective-shift images, video of the captured scene, and stereo image pairs of the captured scene. In addition, there is a facility for importing any manually edited depth maps that can be applied to the Lytro images. The LF images are stored in .lfp and .lfr formats for the Lytro-I generation and Illum cameras respectively.

### **5.2.1. Performance of the Lytro cameras**

The Lytro cameras are built to perform in indoor and outdoor environments to capture LF data with a single operation. The light radiance captured by the Lytro cameras are not affected by bright sunlight or by using multiple Lytro cameras for recording LF data when compared with some 3D measuring devices used in engineering applications such as the Kinect [94], i.e. measurements using time-of-flight are limited under bright light

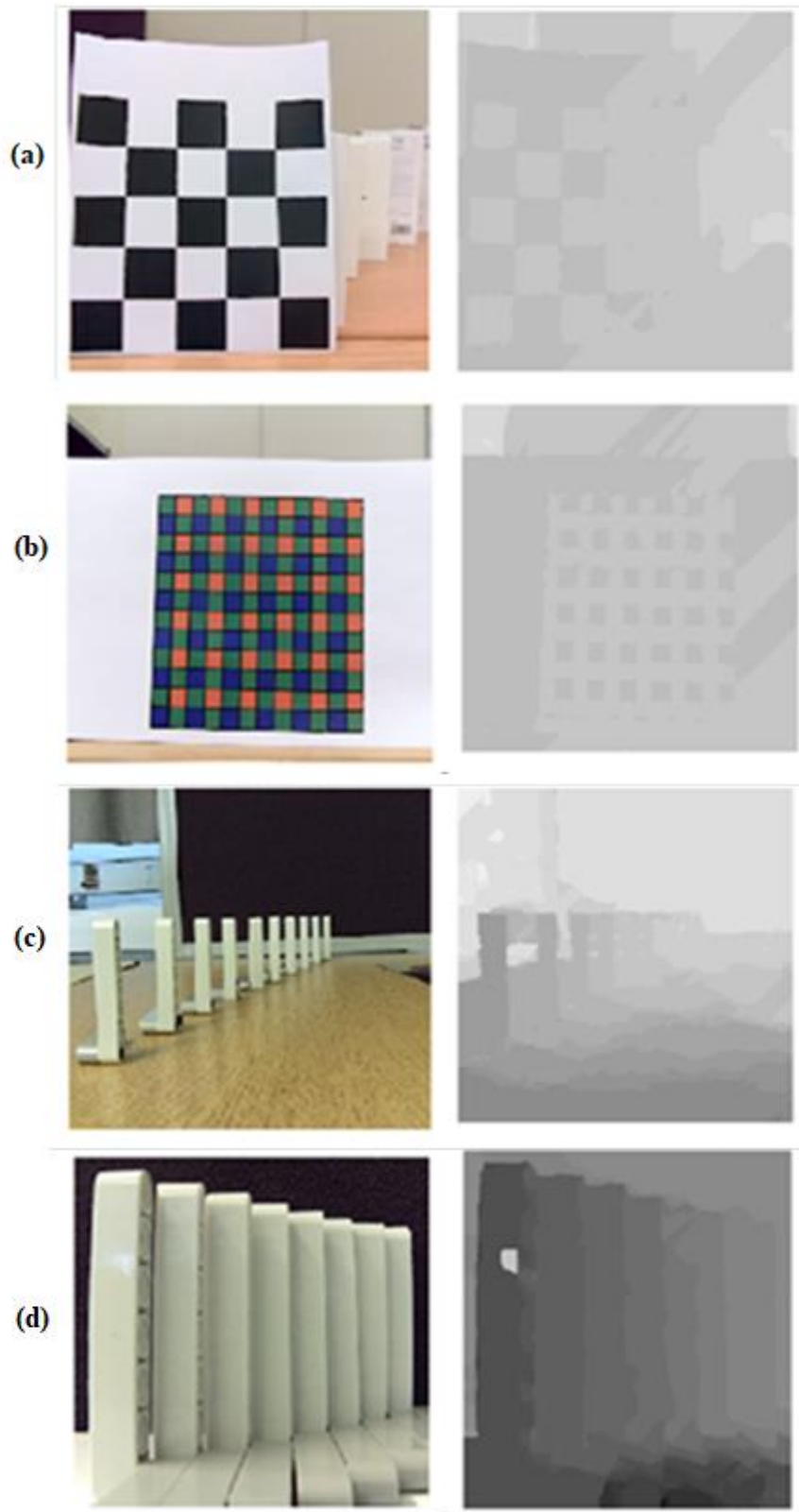


Figure 5.1 Depth response of Lytro camera to (a) Black-White checkerboard, (b) RGB checkerboard, (c) Lego™ bricks placed at different distances to each other and far away from the camera, and (d) Lego™ bricks placed in a compact fashion and close to the camera.

or while using multiple devices simultaneously. This property of the Lytro cameras enables the user to select any environment to capture the LF data. The LDS generates greyscale [0-255, 8-bit] depth maps from the recorded LF data. The Lytro depth results have different response characteristics to colours [95]. This response can be seen especially for the combinations of colours with high-low contrast, for example, white and black, white and red.

This response of the Lytro camera is demonstrated in Figure 5.1 representing RGB scene data (left) and corresponding greyscale depth results(right). The camera generated a depth map of a black and white checkerboard is defined at different depths as shown in Figure 5.1a, where a uniform depth was expected. Similar results can be seen with the RGB checkerboard image (Figure 5.1b), where the blue and green coloured areas are shown at different depths when compared with red alone. However, when using similar or uniform colour objects as shown in Figure 5.1c and Figure 5.1d, the camera can distinguish between several similar Lego™ bricks and each brick is represented with different depth values (a dark and white region in depth map represents near and far regions respectively).

This feature of the Lytro camera potentially makes it interesting to explore for technical applications especially 3D coordinate metrology. It should be noted that in some cases the depth values change for the same bricks when very close to the camera under constant lighting conditions when repeated measurements are performed. This behaviour of the Lytro camera illustrates the potential effects of external factors on depth values. However, the depth values remain consistent when images are taken with constant lighting conditions at a fixed distance from the camera. This provides possible evidence for repeatability, which is a basic requirement of any metrology instrument and devices in engineering applications. This property of the Lytro camera is potentially interesting and can be used in an application where environmental conditions can be monitored and controlled.

A survey of the literature to find any existing computer vision groups using the Lytro camera in engineering applications limited results; with research limited to the use of relative depth [30, 33], and a compression of the Lytro depth map parameters [96] with other techniques. However, no data results are available for generating absolute depth from the Lytro cameras. So, in this thesis, the Lytro depth maps are used with

metrological applications in mind by evaluating camera performance and features in a controlled environment. A relationship between the input data and the output depth map is analysed and the resulting relationship is used to generate the depth data in absolute scale. Also, this research has specifically considered identifying if a working measurement volume can be defined in SI units for a Lytro camera, with initial statements of accuracy and repeatability.

### **5.2.2. Lytro software depth sensitivity to optical distortion**

The Lytro camera is an optical instrument designed to capture light rays along with incoming light direction. Like conventional cameras, both the Lytro-I generation and the Illum cameras suffer from optical distortion (radial and tangential), but LDS has inbuilt process feature to reduce any distortion before generating results (the methods are not disclosed by the Lytro company). However, when objects are placed close to cameras within 100 mm, the distortion is highly noticeable in Illum cameras compared to the Lytro-I generation camera, since incoming light rays are bent unevenly with respect to the optical centre of the camera. The effects of distortion on depth maps and how the Lytro cameras can be potentially used for engineering applications without serious modification to the camera or the results from Lytro software are defined in this section. A key element of this work is to describe some of the procedures to be carried out to use depth maps from the Lytro family of cameras for applications where measuring volumetric distance is the key measurand. Two Lytro-I generation cameras and an Illum camera were used for the experiments in this chapter. This allowed comparison of results between the Lytro-I generation cameras with the Illum camera, as well as a one-on-one comparison of the Lytro-I generation to see how cameras of the Lytro family perform under constant experiment conditions.

The Lytro-I generation cameras and the Illum camera were used to record a flat plane object at 20 mm distance from the camera and generate greyscale depth values using the LDS under a controlled environment with the illumination of 1400 cd. The LF data recorded using the Lytro-I generation camera results in a uniform plane depth map shown in Figure 5.2, whereas depth results generated by the Illum resemble a pincushion optical distortion shape shown in Figure 5.3. The significant depth sensitivity to distortion is because the Illum camera has more microlens groups and each of them can accommodate 96 more pixels when compared to Lytro-I generation

cameras (accommodate 100 pixels). Thus the Illum camera can record 48% more microlens data from the scene and hence the depth values generated by the Illum LF data will be more sensitive to the change in depth than the Lytro-I camera.

In addition, the lens system used in the Illum cameras is designed with multiple lens elements when compared with the Lytro-I generation camera, hence the Illum can observe smaller changes in the scene with respect to the Z direction in front of the camera. More information about this behaviour of both the Lytro cameras is explained in Chapter 6. Due to sensitivity, the effect of radial distortion is clearly visible in the Illum depth results than Lytro I-generation cameras

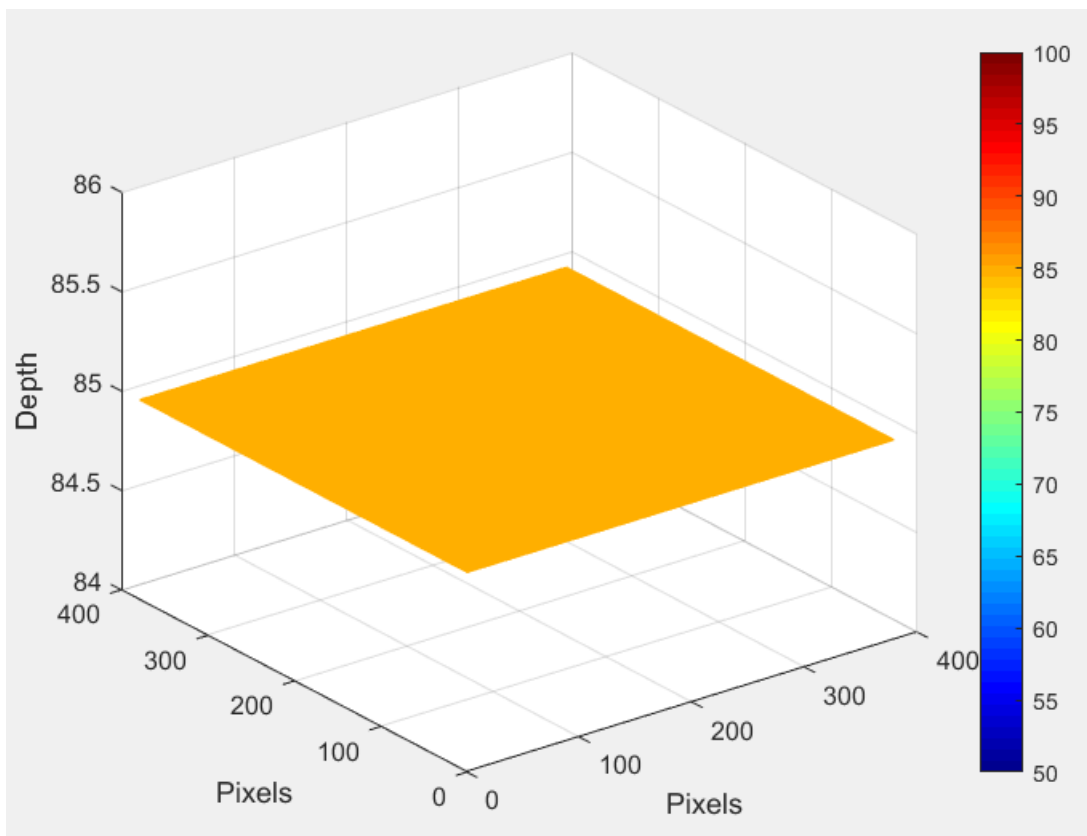


Figure 5.2 The optical distortion effect on depth map for the Lytro-I generation cameras 2D view (top) and 3D view (bottom) for a flat surface at 20 mm distance from the camera

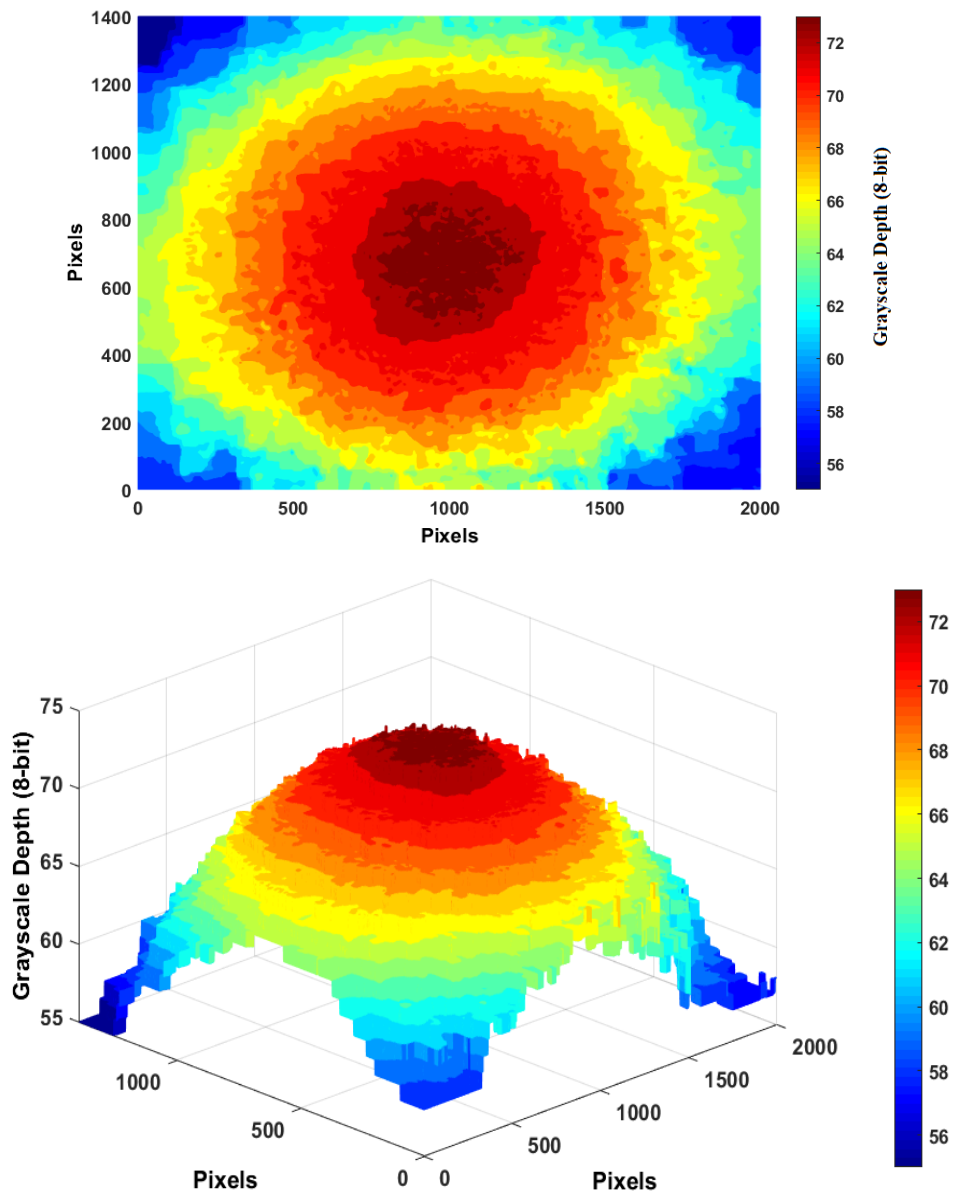


Figure 5.3 The optical distortion effect on a depth map of the Illum camera 2D view (top) and 3D view (bottom) for a flat surface at 20 mm distance from the camera

To use the depth results of the Illum camera without major alterations to the camera or the software, an alternative approach of average depth mapping was used in this work. This method was followed because the distortion effects are highly noticeable when objects are close to the camera, for example highly noticeable distortion of checkerboard image looks less distorted when far from the camera, but distortion values from the centre of the camera remain the same. So, using negative map values calculated using depth images captured when objects are close to the camera, will not alter the original depth results of faraway distance objects.

To calculate the depth value using the averaging technique, the optical centre of each image was found by calculating the centre pixels of the output image, with  $(c_x, c_y)$  being the optical centre of the image,  $r$  the distance of image pixel  $(x, y)$  under consideration from the optical centre (Equation 5.1), and  $r_c$  the radius under consideration that decreases as the fixed flat object moves away from the camera (Equation 5.2), as shown in Figure 5.4 (top). Since the Illum raw image width is 1.4 times the height, there are two reasons to choose  $r_c$  along the width of the image instead of the height of the image. Firstly, since the distortion was increasing outwards from the optical centre, it is better to consider maximum depth values of the surface for averaging. Secondly, if the radius was considered along height, the total length of  $r_c$  will be less than the actual length of the object and hence total points for averaging will be less. Using Equation 5.3, the average depth value for a flat surface at a given distance from the camera can be calculated with  $N$  being the number of points within  $r_c$  radius, while  $m, n$  are the pixel count of the photosensor in  $x$  and  $y$  direction respectively .

$$r = [width/2] \quad 5.1$$

$$r_c = \sqrt{[(C_x - x)^2 - (C_y - y)^2]} \quad 5.2$$

$$d_{Avg} = \begin{cases} \left( \sum_{x=0}^m \sum_{y=0}^n [I(x, y)] \right) / N & \in r_c \leq |r| \\ 0 & \in r_c > |r| \end{cases} \quad 5.3$$

This method of averaging is applicable for the Illum camera because the rate of distortion is independent of the object distance from the camera. A value of  $\bar{n} = 5$  has been used for all the experiments and the final depth values were calculated using Equation 5.4, where  $d_i$  is the greyscale depth values (for the Illum camera  $d_{Avg} = d_i$ ),  $n$  is the number of images and the resulting average depth map is shown in Figure 5.4 (bottom).

$$d_{final} = \left( \sum_{i=1}^{\bar{n}} d_i \right) / \bar{n} \quad 5.4$$

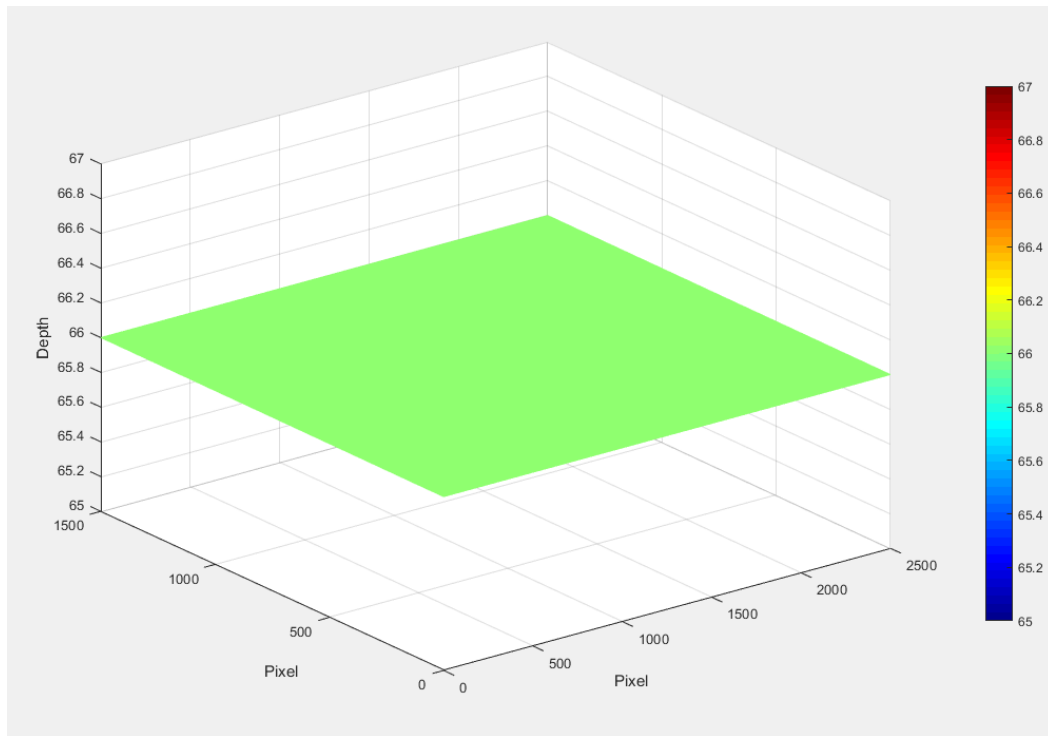
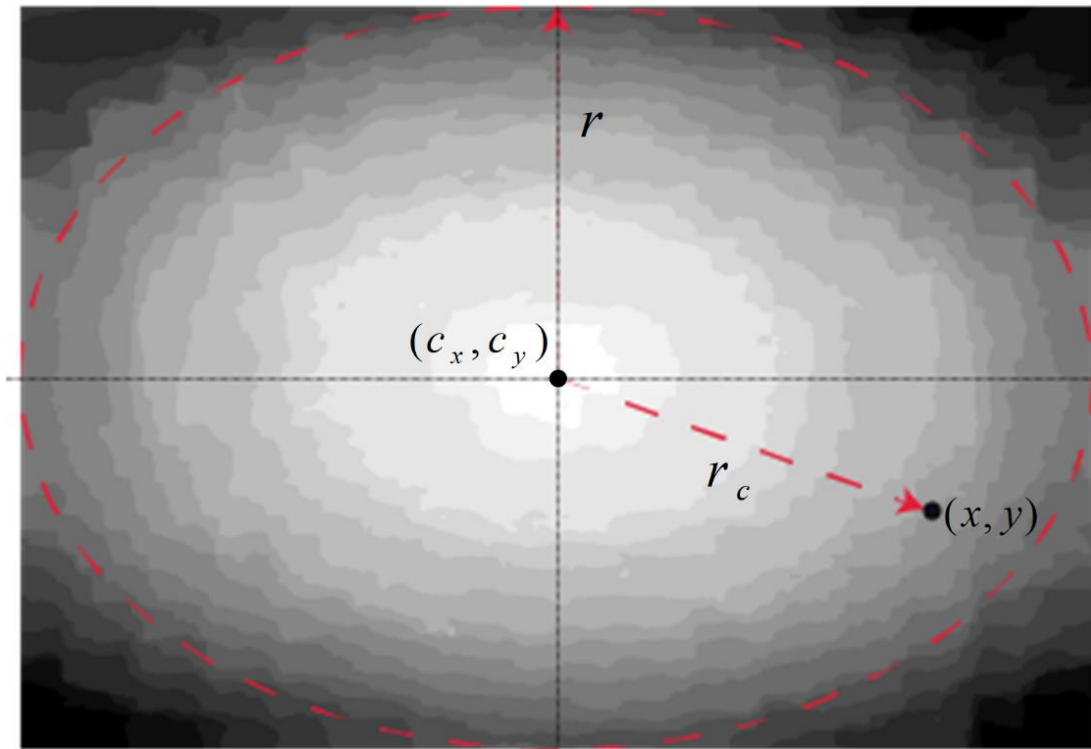


Figure 5.4 Represents region selected for averaging technique (top) and 3D view of distorted depth correction result (bottom)



### **5.3. Depth Calibration setup**

The calibration setup was achieved by having a stable laboratory environment under controlled lighting (typically 1400cd), humidity (50%±10%) and temperature (20°C ± 0.5°C) conditions for all experiments. The main theme of the work was to measure distance related depth values which range from 0 to 255 for the 8-bit depth map data. The Lytro cameras were in the same position and orientation, for all changes with object distance. Accurate positional information of distance between the camera and object was recorded using a motorised positional encoder (linear rail) unit with an accuracy of ±10 µm (as defined by Renishaw laser interferometer XL-80). A uniform diffuse lighting system was used to avoid directional and specular light entering camera objective from the object surface. In this experiment, the camera image plane and the object plane were set to be parallel with respect to each other, to obtain a uniform depth value for any given distance from the camera as shown in Figure 5.5.

Five images of the flat object surface were recorded. All five images were then processed through LDS to generate an 8-bit depth map. These depth maps were then used to generate the average greyscale value for a given distance. The same procedure was followed until the object reached 1000 mm, with 5 mm constant increment in object distance from the camera.

### **5.4. Specification of the Lytro cameras, software and external factors**

Two Lytro-I generation cameras (LC1 and LC2) and an Illum camera (LC3) were used in the experimentation to determine Lytro cameras depth response to changing distances. The LC1 and LC2 have a slight variation in microlens array alignment of 0.35° and -0.5° respectively with respect to the photosensor plane, while LC3 has a 0.0° alignment (Table 5-1)(see Chapter 3, sec 3.5) (see Appendix 10.1 and 10.2). All other features on these cameras were set to be in the automatic mode, with shutter speed, ISO (sensor sensitivity of photosensor) [97], and neutral density filter values noted as being 1/40 of a second, 400, and -0.4 respectively, at 120 mm object distance from the camera. The Lytro Desktop application of version 5.0.1 was used to generate all-in-focus images and depth map images in the experiments.

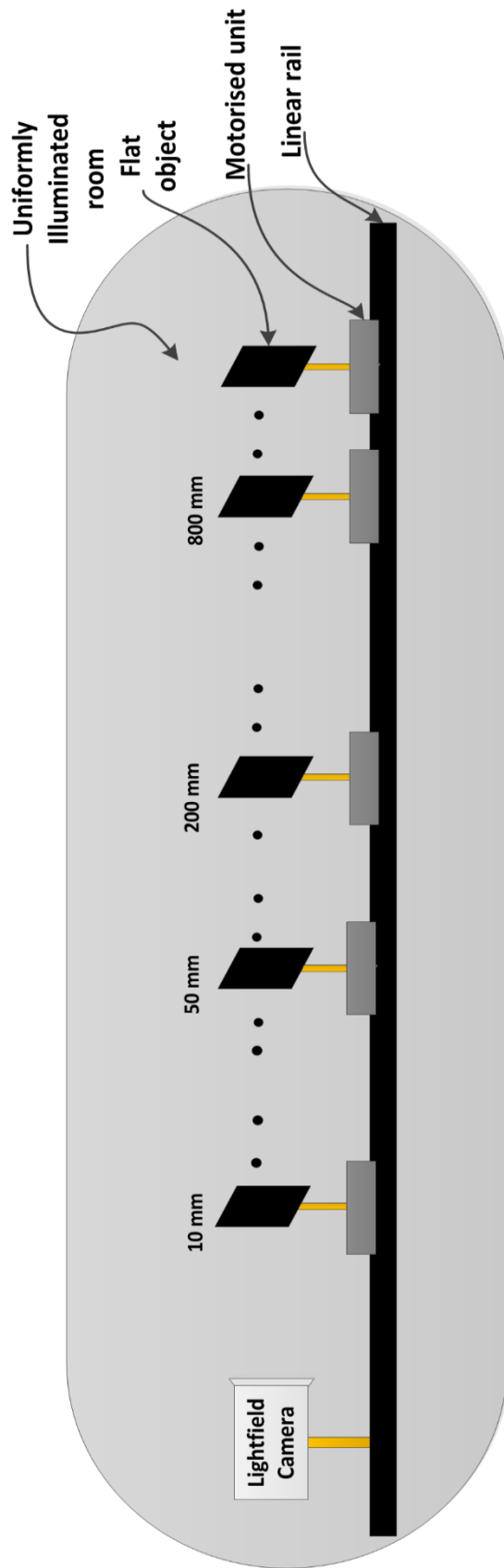


Figure 5.5 Experiment set-up to generate distance related greyscale values

Table 5-1 Specifications of the Lytro family of cameras used in-depth map calibration experiment

<b>Features</b>	<b>LC1</b>	<b>LC2</b>	<b>LC3</b>
MLA	376 x 327	376 x 327	434 x 541
MLA rotational error (degree)	0.35	-0.5	0.0

All experiments were carried out, one camera at a time, by varying only the object distance from the camera under constant lighting conditions (1400 cd or 1600 cd) and using the same object. The non-reflective diffused paper was attached to a flat glass plate (200 mm in height and 150 mm in width) and used as the object for all experiments. The non-reflective face of the object was made to face towards the camera and five images were captured for every 5 mm increment in the object distance from the camera up to 1,000 mm. These Lytro images were then processed using the LDS tool and MATLAB 2015b to obtain the final average depth map as shown in Figure 5.6. Both RGB image and depth maps were generated for a group of images using the LDS tool.

Using the RGB image, the pixel coordinates of the object under consideration was identified and the same pixel coordinates were used to select the object space in the depth map. Since the RGB and the depth map generated by the LDS tool have the same pixel count (width and height), matching the object space in the depth map was made straightforward. After averaging each pixel in all five depth maps that were matched with the object under consideration, a new greyscale depth value was generated for a distance and this procedure was carried for all groups of images. Thereby a new averaged depth map was generated for every 5 mm increment in distance to be used later for plotting sensitivity curves.

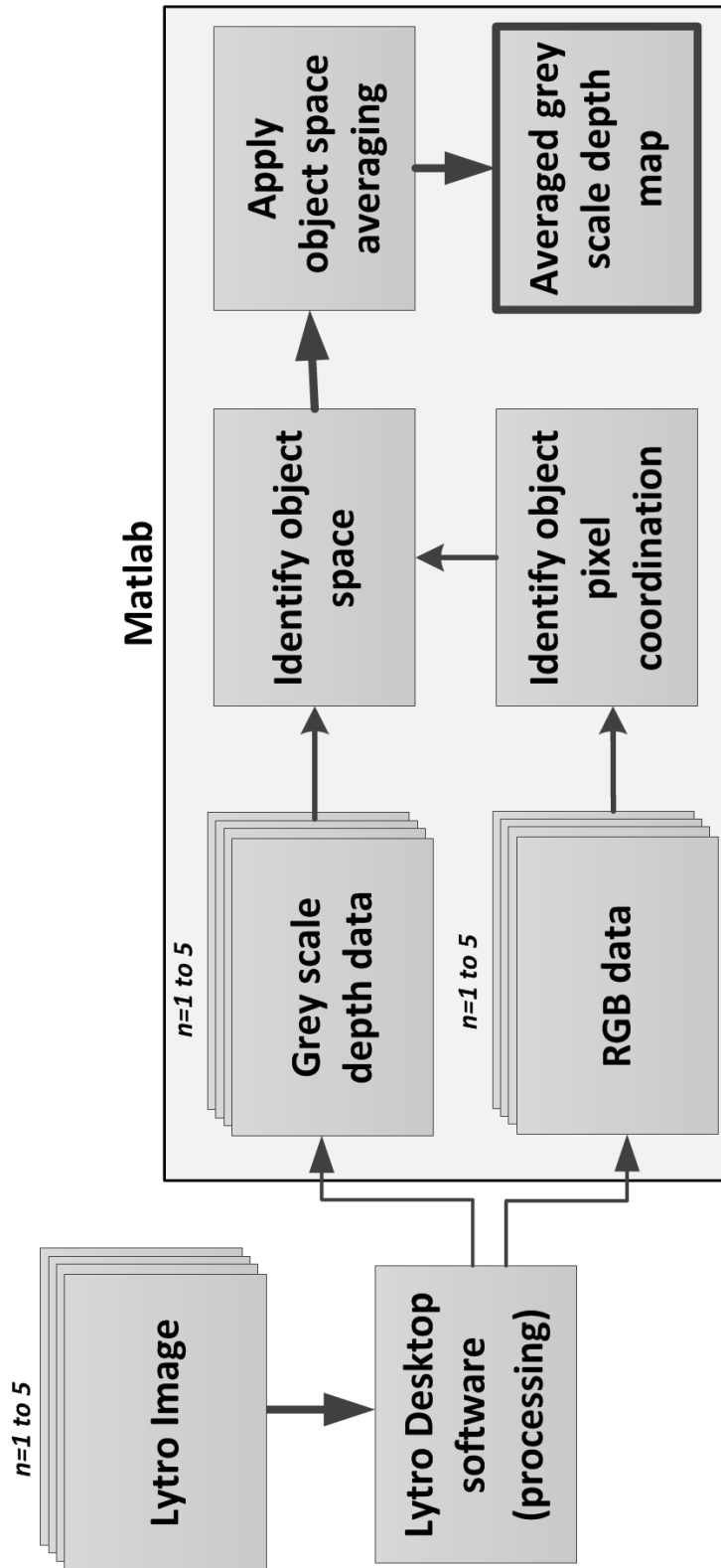


Figure 5.6 Procedure to generate an average greyscale depth map

## 5.5. Response curve

Depth range and repeatability are among the important features of the depth measuring instruments that are used in metrological applications. Depth range defines the distance from an instrument that can be measured, and results are ideally required to be highly repeatable. Repeatability defines the ability of the instrument to produce the same results measured in different instances, provided that there is a constant measurement condition. With a large depth range and repeatability value, the measuring range and the confidence of using measured data in any application will be high by a potential user. Similar to all depth sensing devices, the Lytro cameras have a defined depth range which is potentially influenced by lighting conditions and surface nature of the object measured. In this work, this novel development has been called the response curve. The response curve is a method of expressing the relationship of how depth values measured using a Lytro camera can be expressed in SI units. It also represents the camera behaviour for a given working environment (lighting condition set up).

The response curve results for the Lytro family of cameras are not linear. Depth values vary with distance and gradually reach a constant value as shown in Figure 5.7 (it should be noted that multiple responses have been presented in each case). The results indicate that the entire response curve from the Lytro cameras cannot be used as results for applications. This suggests that only a section of the response curve gives viable and useful data that has a linear relationship with the greyscale depth to absolute distance (m). All three response curves have been divided into an Active Zone (*AZ*) and Inactive Zones (*IAZ*) depending upon the relationship between both axes, greyscale depth values and distance from the camera. An *AZ* is categorised as a region where the resulting response curve has a linear variation with respect to both axes, while an *IAZ* is a region where there is no possible linear relationship between two axes (distance to grey scale values), i.e, large distances are represented by a small number of greyscale values (very low sensitivity response) or a large number of greyscale values are represented by small distance values (very high sensitivity response). Standard deviation typically of  $\pm 5$  for a grayscale value can be plotted on the graph but due to scaling issue these are not observable. Hence in Figure 5.7, the error bar have not be included.

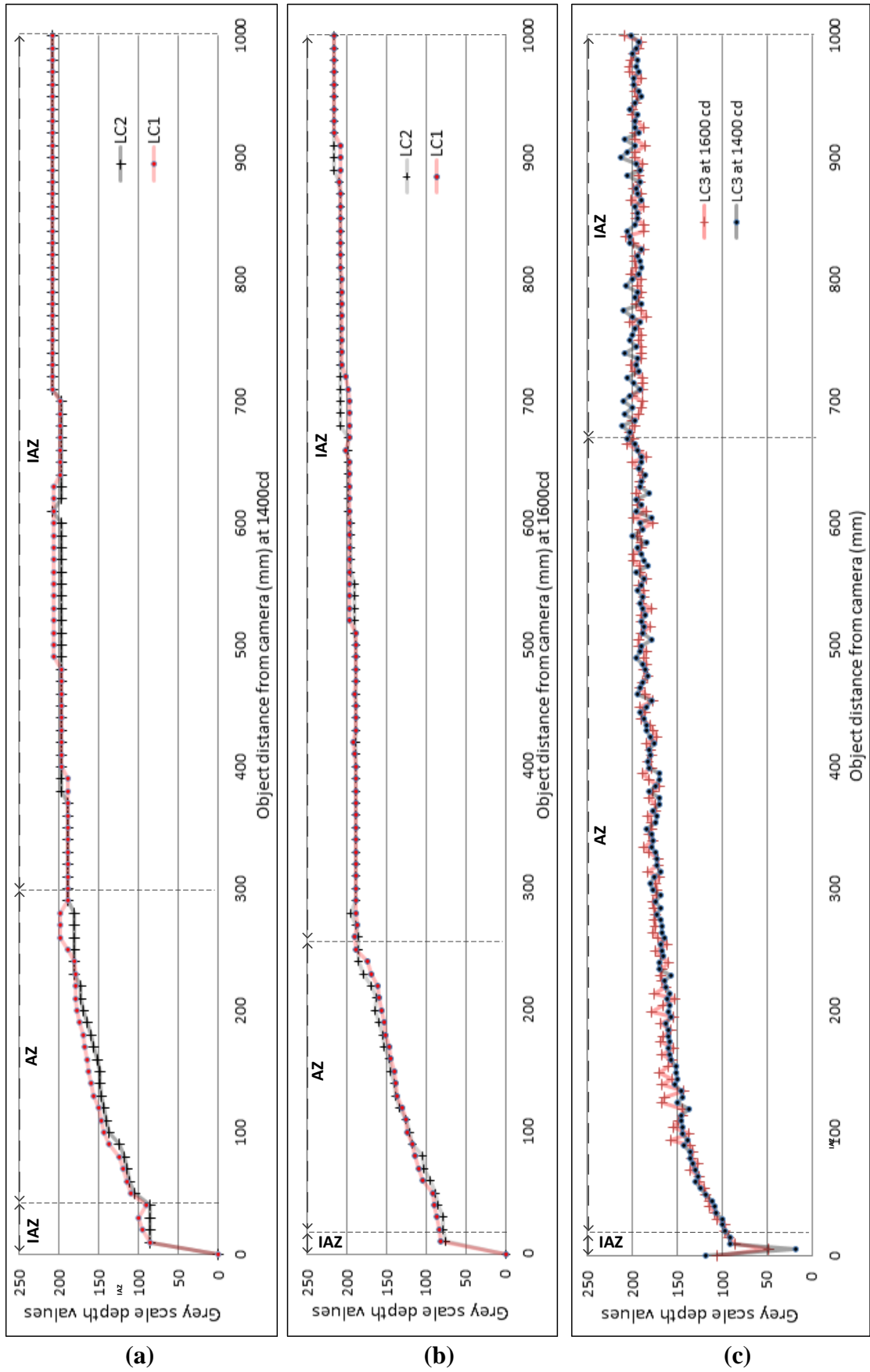


Figure 5.7 Response curves of LC1 and LC2 camera at 1400cd (A) and 1600cd (B) respectively, and LC3 camera (C)

The LC1 and LC2 cameras (Figure 5.7A and Figure 5.7B) have a very limited initial response for close range depth detection because 0 mm is represented by 0 greyscale depth values and later on rapidly increases to 85 greyscale depth values for 10 mm measurement on both cameras. A similar trend is shown by the LC3 camera (Figure 5.7C) where close range measurement starts with 130 greyscale value, drops to 25 greyscale value and then increases steadily.

The depth values from 0 mm to approximately 40 mm and after 280 mm, are defined as *IAZ*, while 50 mm to 270 mm is defined as *AZ* for LC1 and LC2 cameras. For the LC3 camera, the *AZ* stretches from 10 mm to 500 mm, after that the response is constant to change in the distance which is therefore defined *IAZ*. From Figure 5.7C, it appears as if the response is changing after 500 mm, but when averaging is applied to LC3 response curve, the response becomes constant after 500 mm. Also, it can be seen from the LC3 response curve that after 500 mm even the variations become constantly repeating.

The response curve for all three Lytro cameras represents the sensitivity of each camera to a given working environment, where lighting conditions and object surface have great influence on the shape of the response curve. The *AZ* for each Lytro camera can be used as an indicator to understand the working distance of the camera so that the user can obtain viable results which can be used in practical applications with confidence. Also, the *IAZ* is an important term to know since this defines the starting point of the response curve where unreliable sensitivity phase begins. Having defined the *IAZ* for each camera under fixed conditions, the user is then informed of the fact that the camera should not be used in any application where the required working distance extends into the *IAZ*.

## **5.6 Pixel resolution of the Lytro cameras**

The *Z* axis measurement of a given object or scene is very important in metrological applications, but for complete 3D analysis of an object or scene, it is also very important to measure the *X* and *Y* coordinates as well. For machine vision applications, it is important to calculate the height and width of an object by counting the number of pixels accommodated in the region of interest (ROI) and multiplying the number of pixels with the corresponding pixel resolution providing the actual measurement. For such calculation, the final image from the machine vision camera should be free from distortions and blur.

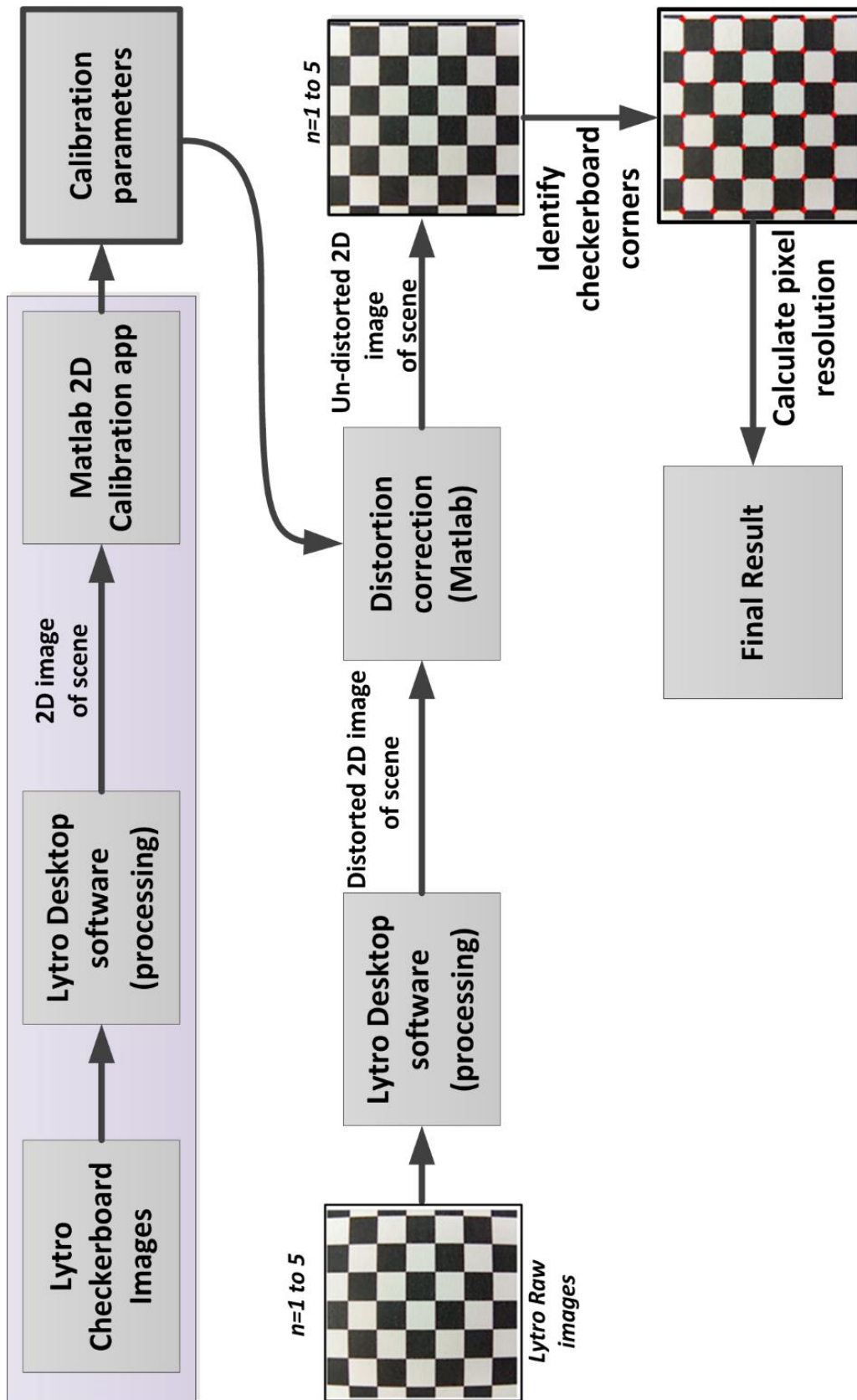


Figure 5.8 Procedure to calculate pixel resolution



Optical distortion will reduce the accuracy of any measurement application using the camera by disrupting the final image by barrel distortion or pincushion distortion or mustache distortion effects [98]. Also, blur causes problems in identifying the exact number of pixels in the ROI. One of the advantages of using the LF camera for measuring spatial resolution is that along with the greyscale depth map, the software generates an all-in-focus RGB image of the scene. The all-in-focus image has all depths from the camera at higher focus values and hence blur-free RGB image.

Similar to other digital cameras used in machine vision, the Lytro cameras also suffer from optical distortion. This distortion will again pose a problem for calculating the exact number of pixels in the ROI, so it is important to correct the optical distortion before calculating each pixel value in metric units (mm). The Lytro camera records scene data in 4D but using the LDS tool it is possible to extract the RGB all-in-focus (2D) image and this image is used for spatial calculations. The Lytro RGB image was processed to generate radial ( $K1, K2$  and  $K3$ ) and tangential ( $P1$  and  $P2$ ) distortion parameters using MATLAB distortion code. These parameters are later used to generate distortion-free images for further calculations (see Table 5-2). The procedures followed to generate the Lytro calibration parameters and using it to generate distortion-free Lytro image are shown in Figure 5.8. A group of 15 images of a regular pattern checkerboard was captured using the Lytro camera and these images were fed into the calibration algorithm (15 images because, the calibration algorithm requires around 12 to 15 images for efficient functioning). The algorithm processed the images and generated the calibration parameters for a given camera. The same procedure was applied to all three Lytro cameras and individual camera parameters are stored. Once the calibration parameters were generated, any 2D images extracted from a Lytro camera were processed using the MATLAB distortion correction function to obtain a distortion-free image. Note that whilst optical lens correction is achieved as defined in Figure 5.1, the residual analysis from true geometry cannot be accessed, because the LF cameras are interpreting grey scale values at different depths even for a flat plane object.

Table 5-2 Calibration parameters of Lytro family of cameras

<b>Camera</b>	<b>K1(pixels)</b>	<b>K2(pixels)</b>	<b>K3(pixels)</b>	<b>P1(pixels)</b>	<b>P2(pixels)</b>
<b>LC1</b>	-0.1980	0.3640	-1.5740	-0.0013	-0.0021
<b>LC2</b>	-0.1583	0.1111	-0.1788	-0.0011	-0.0032
<b>LC3</b>	-0.0291	0.1051	-0.2067	0.0022	0.0009

To calculate each pixel value in SI units (mm), a regular pattern checkerboard (white-black of 25.1mm grid size) was used as an object, placed in front of the Lytro cameras. Raw images of the checkerboard were taken at different distances, ranging up to 1,000 mm with an increment of 50 mm in steps from 0 mm. For each increment in distance 5 images were taken and processed for pixel resolution. Using pre-calculated calibration parameters on distorted Lytro images, a distortion-free image was generated, and checkerboard corners were detected. Furthermore, the number of pixels in between the checkerboard corners were calculated for each image. For a given distance from the camera, the same procedure was followed on all 5 images and the resulting pixel count was averaged to obtain the final pixel count. This value must be normalised to obtain an SI pixel value (mm), so the final pixel count was divided by the actual dimensions of the checkerboard. Let  $(c_h, c_w)$  be the height and width of the checkerboard per unit,  $d$  be the distance of the checkerboard from the camera and  $(n_h, n_b)$  be the average pixel count in each unit of the checkerboard as given by Equation 5.5 and Equation 5.6. The final pixel resolution for all three Lytro cameras is shown in Figure 5.9.

In Figure 5.9, the pixel resolution of LC1 and LC2 cameras are similar with respect to the change in distance, representing the consistent build quality. So the pixel resolution values calculated in this experiment can potentially be used for all Lytro-I generation cameras since the build quality is very similar but minor alignment issue of MLA with the photosensor may exist. The response curve experiment for LC1 and LC2 cameras also show the close resemblance between both Lytro-I generation cameras.

When comparing the LC3 camera with LC1 and LC2, it can be noticed a slight drop in the pixel resolution value from 300 mm to 1000 mm, suggesting varied behaviour of Illum camera with respect to Lytro-I generation cameras. Since a single Illum camera was used in this experiment, the built quality of Illum cameras and the reason for drop-in pixel resolution have not been investigated in depth.

$$Pixel\ count = (n_h/c_h, n_w/c_w) \quad 5.5$$

$$Pixel\ count_{final} = \left( \sum_{i=1}^n Pixel\ count_i \right) / n \quad 5.6$$

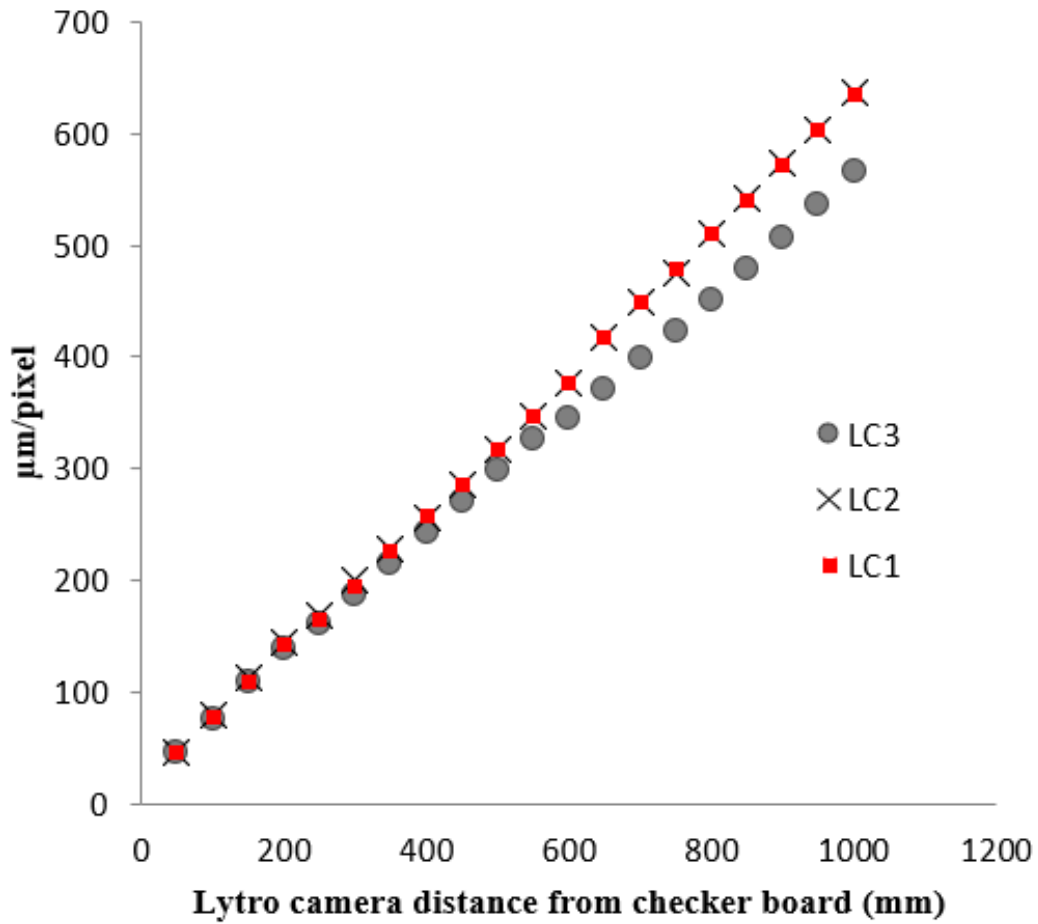


Figure 5.9 Pixel resolution of Lytro cameras

## 5.7 Result 3D Measurement

The response curve data and the pixel resolution data of the Lytro family of cameras are the key information to relate the cameras to their respective measuring capabilities. With this information, it is possible to generate depth maps that contain varying greyscale values into greyscale values representing absolute depth. To accomplish the task of generating 3D depth data in absolute scale, the collected response curve and pixel resolution data has to be used along with the captured scene.

Initially, using the LDS tool the Lytro raw image was processed to generate greyscale depth data and a 2D-RGB colour map of the scene. The greyscale data was then used with response curve data to generate the depth data in absolute SI units (mm). Each greyscale value in the depth map from the Lytro data was matched to the pre-calculated response curve data and corresponding SI value was generated. Thereby, complete SI depth data were obtained, with each Lytro camera having its own Z-axis measurement range. With the help of 2D-RGB data and pixel resolution data, calculating the width and height of objects in the scene was achieved. The number of pixels within the ROI was calculated and the SI depth of the ROI was determined by depth result. The pixel resolution value corresponding to the SI depth data was then used to generate X and Y axes values in SI units. Thus the complete 3D data was generated using data obtained from the Lytro cameras.

Figure 5.10 represents the workflow to generate 3D data from the raw Lytro scene data. Figure 5.11 and Figure 5.12 illustrates the raw image of the Lytro-I generation and the Illum cameras respectively, and the final SI depth data for four objects marked 1 to 4. The 3D validation values are represented in Table 5- and Table 5-. The 'x' notation in Table 5- indicates the occlusion effect introduced due to which measuring total pixel count of target object was not possible. Further, examples of 3D depth data in SI units are represented by Figure 5.13, Figure 5.14 and Figure 5.15. The depth data along with the RGB values obtained by all-in-focus image was rendered using a 3D viewer software package (MeshLab) and the screenshot is represented by Figure 5.16. The left-hand side represents the top view of relative data from the Lytro software, while right-hand side of the figure represents the Z calibrated Lytro depth data. The verification data demonstrates the difference between expected measurements and calibrated Lytro measurements (Table 5-3).

The accuracy and repeatability achieved were approximately +10.0 mm to -20.0 mm, and typically 0.5 mm respectively in Z coordinate (depth). The drop in the accuracy was due to the step size of the response curve (5 mm) and hence accuracy was a multiple of response curve step size. The lateral results match very closely within the range of  $\pm 1.5$  mm and it was observed that the depth and lateral results for different illumination conditions (1400 cd and 1600 cd) are in the range defined earlier.

Table 5-3 Validation of 3D measurement shown in Figure 5.11

Object	Real dimensions (mm)		Real Distance from camera (mm)	Trial			Calculated dimensions (mm)		Calculated distance from camera (mm)
	Height	Width		1	2	3	Height	Width	
1	99.5	98	280	180	181	180	98.2	96.7	290
2	39.9	40.1	240	178	178	178	41.2	42.4	240
3	39	39.5	180	161	160	160	38.6	38.5	175
4	40.1	40	130	132	142	141	39.2	40.6	140

Table 5-4 Validation of 3D measurement shown in Figure 5.12

Object	Real dimensions (mm)		Real Distance from camera (mm)	Trial			Calculated dimensions (mm)		Calculated distance from camera (mm)
	Height	Width		1	2	3	Height	Width	
1	99.5	98	210	180	178	180	x	96	220
2	39.9	40.1	180	161	161	160	41.2	x	180
3	39	39.5	180	161	161	160	38.5	x	175
4	40.1	40	80	122	123	123	39.1	40.5	80

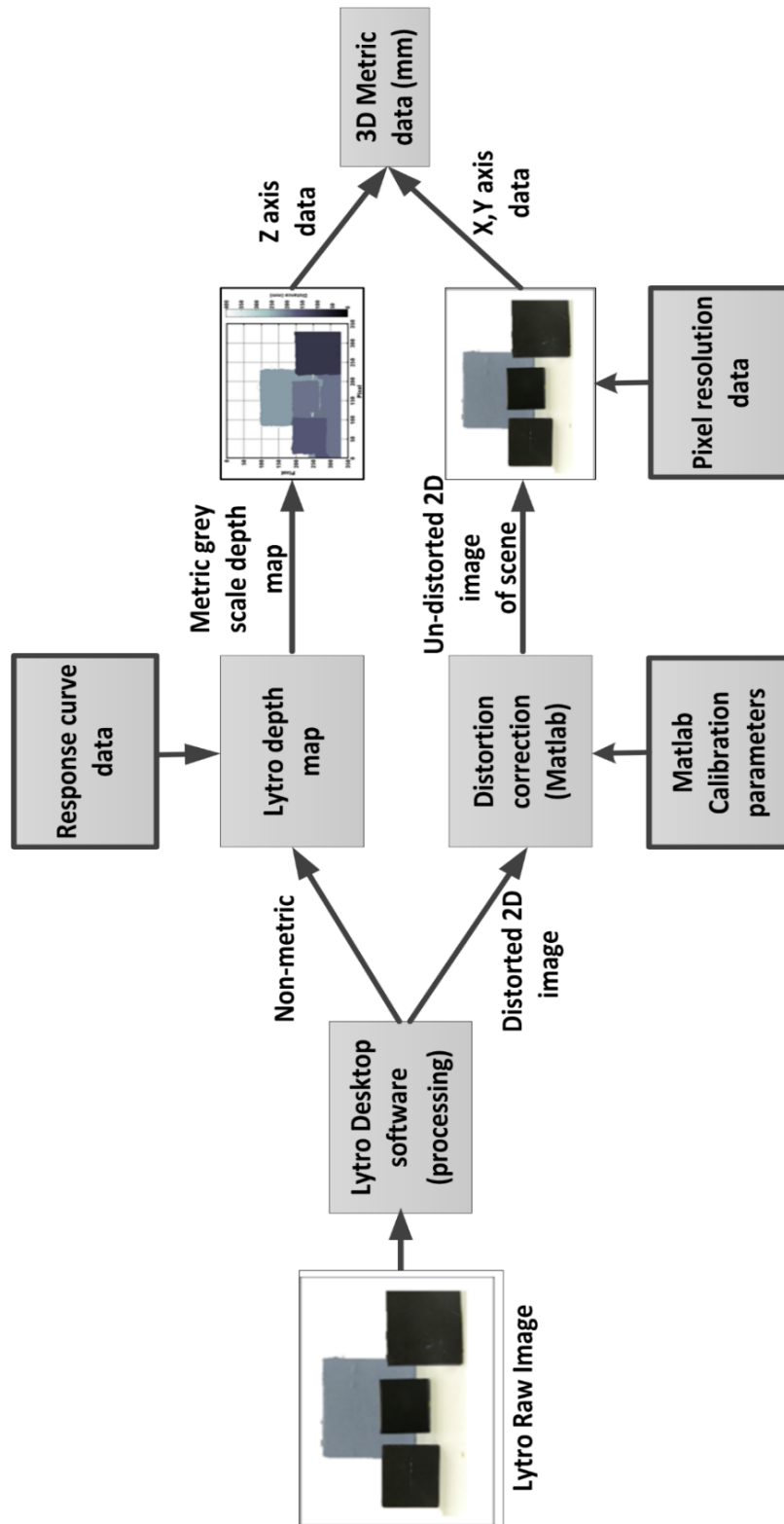


Figure 5.10 procedure to generate 3D data from Lytro raw data

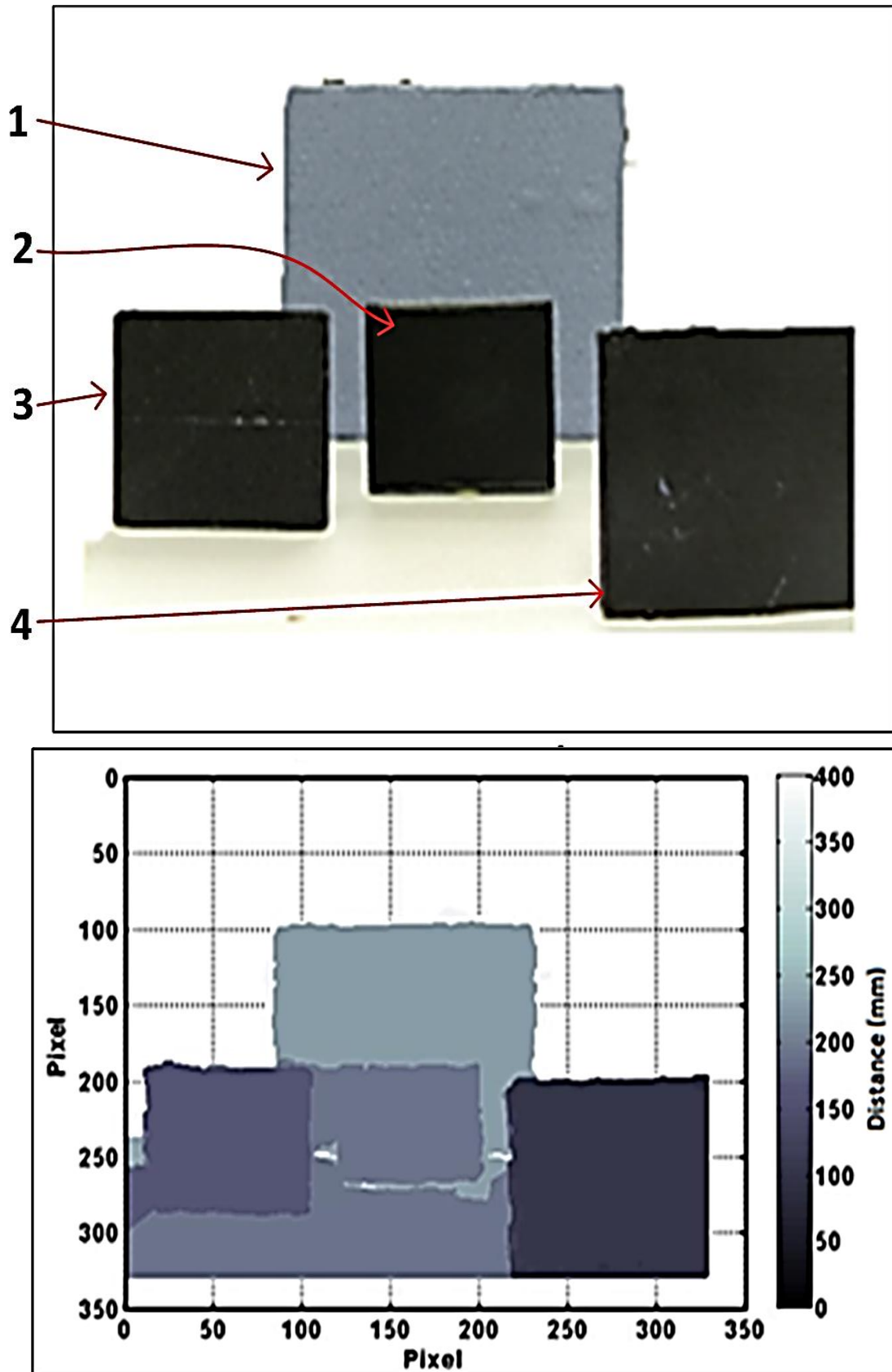


Figure 5.11 Raw scene data (top) and corresponding metric depth data (bottom) of the Lytro-I generation camera

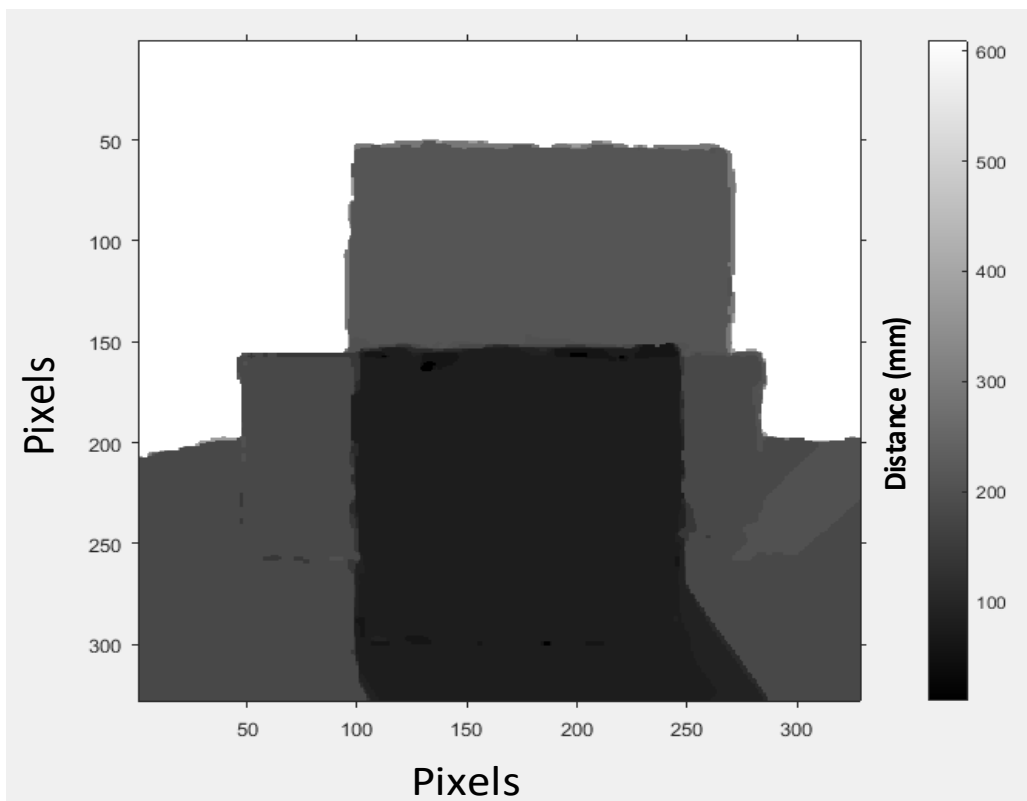
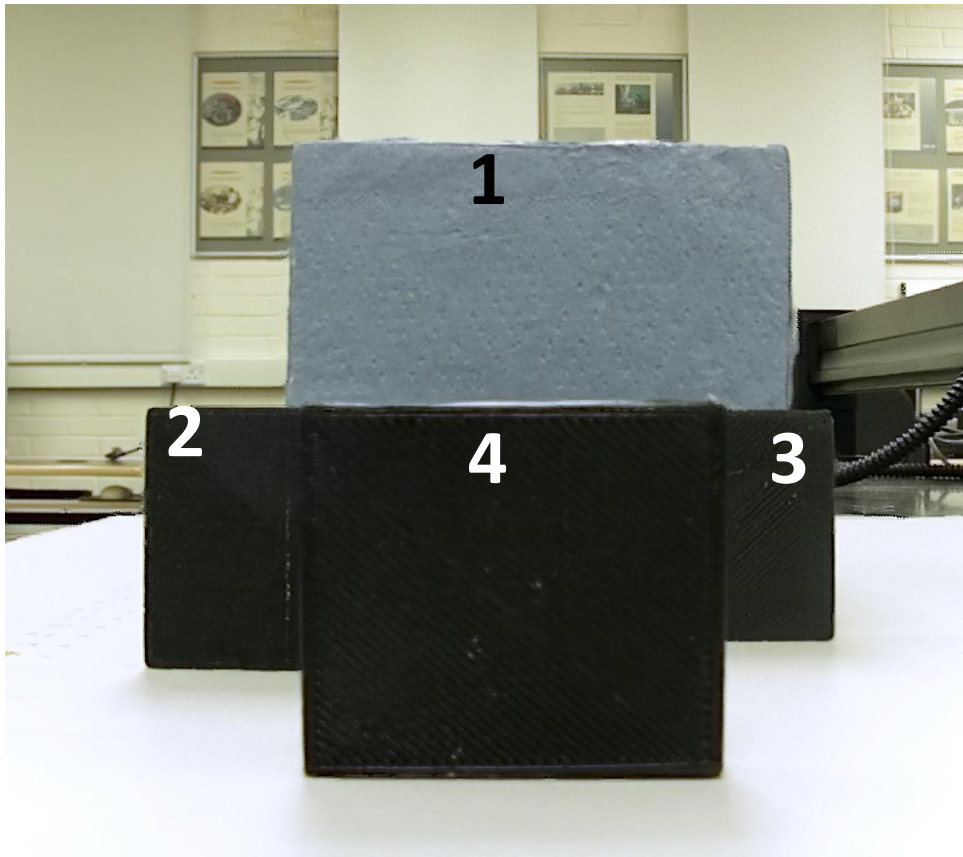


Figure 5.12 Raw scene data (top) and corresponding metric depth data (bottom) of the Lytro Illum camera



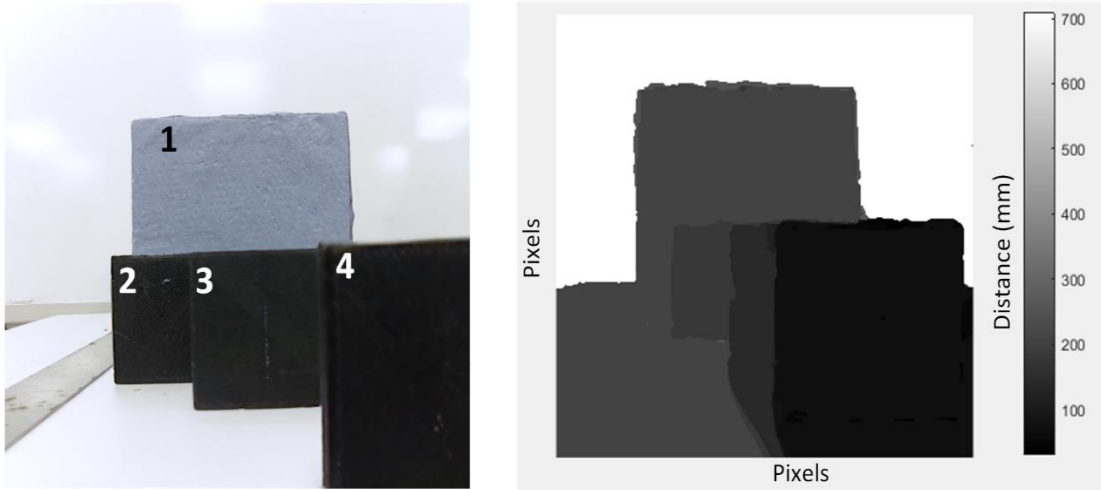


Figure 5.13 Lytro-I generation (LC1) RGB scene (left) and corresponding depth map (mm) (right)

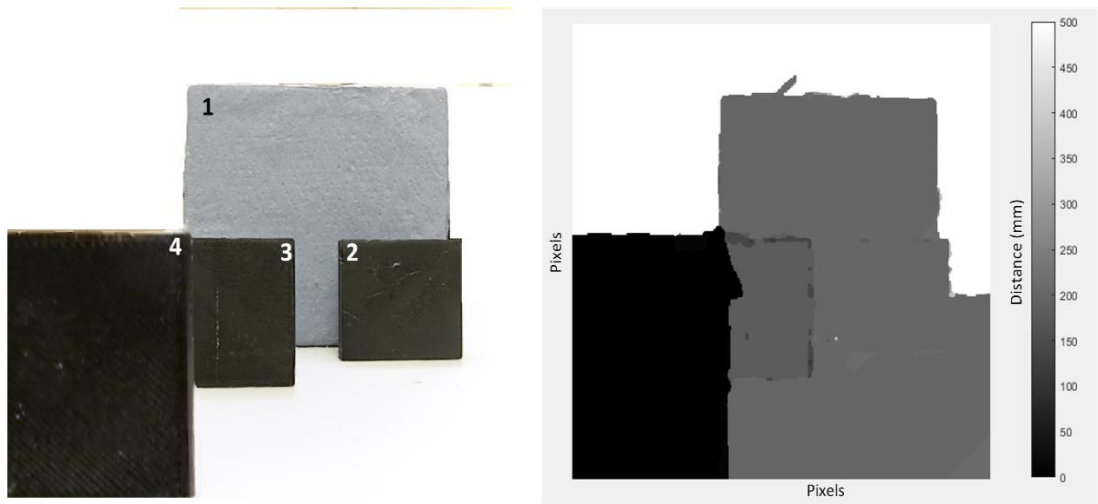


Figure 5.14 Lytro-I generation (LC2) RGB scene (left) and corresponding depth map (mm) (right)

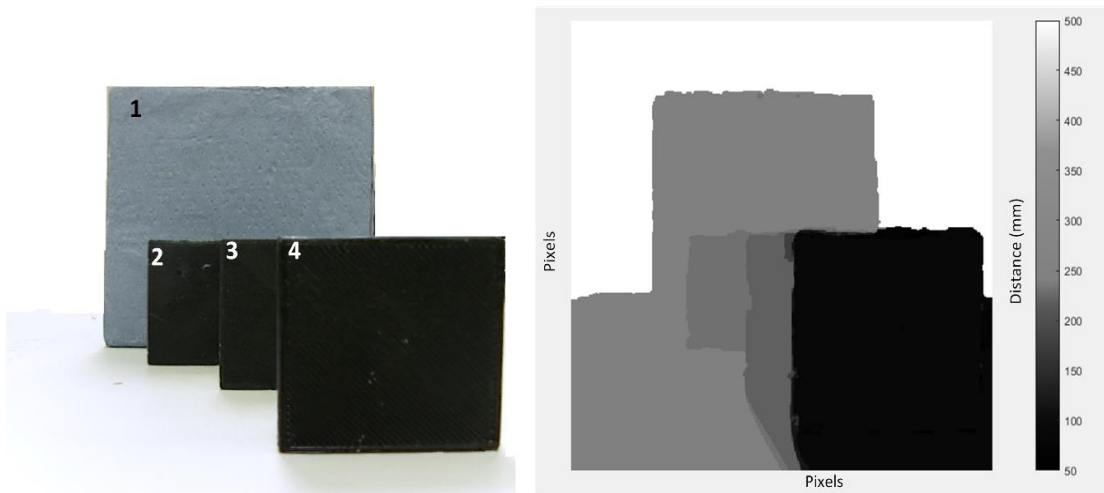


Figure 5.15 Illum camera (LC3) RGB scene (left) and corresponding depth map (mm) (right)

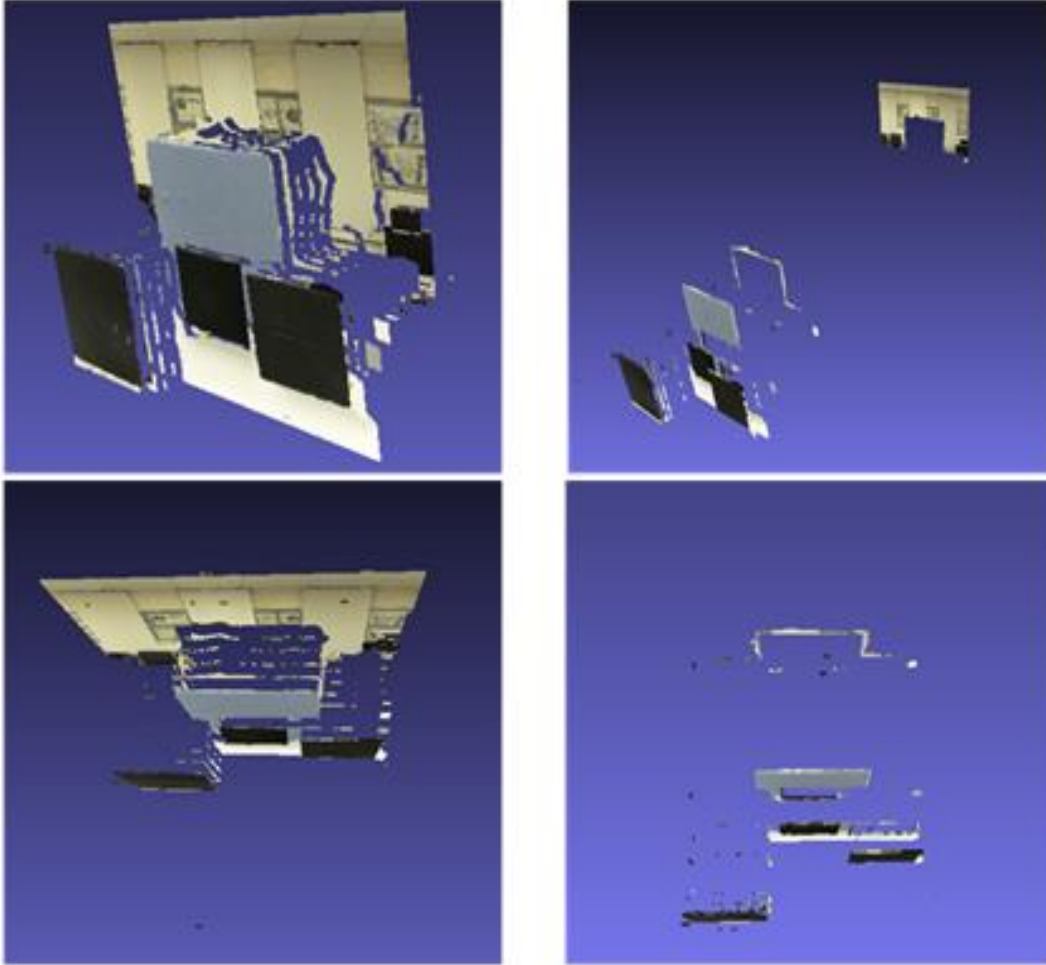


Figure 5.16 A screenshot of depth data rendered along with the RGB value using a 3D viewer software package (MeshLab) (Lytro relative depth result and right- Z calibrated data)

## 5.8 Lytro response to uniform illumination

In earlier experiments carried out with both versions of the Lytro cameras, the objects were placed in a controlled environment parallel to the LF cameras. Hence, the cameras were forced to capture the reflected light from the object surface. Thereby, suggesting that depth maps depend on the light intensity/illumination.

Furthermore, to understand the behaviour of depth data generated by the Lytro family of cameras for a uniform illumination emitted from the object surface, experiments were conducted with the help of a laptop screen as an object. The cameras were placed parallel to laptop screen with user-controlled brightness and colour patterns made visible on the screen. The images of different colours were displayed on the laptop screen and captured using the Lytro family of cameras. For a different combination of

colours, five images of the laptop screen were taken, and a uniform depth map was expected from Lytro software. Figure 5.17 represents the outcome of uniform illumination experiments under different colours scenario and corresponding RGB values and greyscale depth values are shown in Table 5-4.

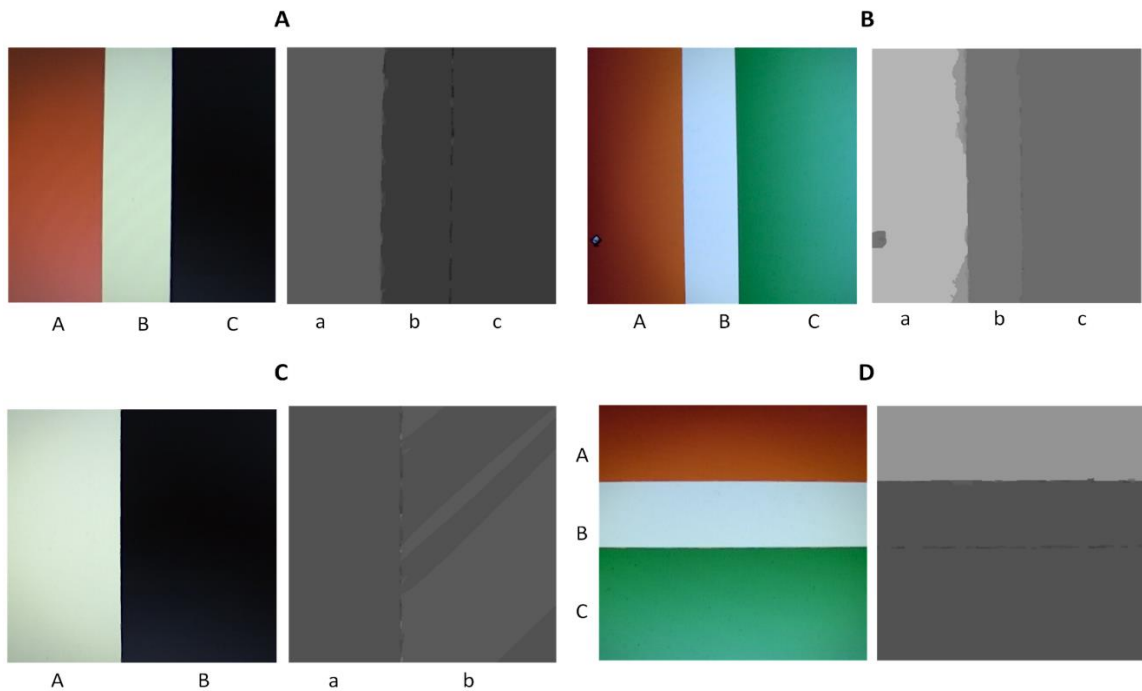


Figure 5.17 Results of uniform illuminations with different colour and corresponding depth maps from the Lytro family of cameras. The top row represents images of the Lytro-I generation camera and bottom row represent Illum camera

Results shown in Figure 5.17 illustrate the nature of the LDS (black box) in generating the depth maps. With uniform illumination, it was observed in the depth maps that different regions (greyscale values) were representing the uniform depth of the laptop screen, where uniform depth was expected. So, the LDS does not just depend on the light intensity to generate a depth map but also has some other independent and additional variable on which nature of resulting depth map depends. This behaviour of the Lytro camera indicates the need for understanding black box software or to have a working software model of which all variables are known and can be independently controlled.

Table 5-4 RGB and Greyscale values of images used in the experiment shown in Figure 5.17

Image	RGB value	Depth Greyscale values
A	A [ 166 74 44]	a [90]
	B [201 225 204]	b [58]
	C [10 10 11]	c [58]
B	A [155 74 37]	a [181]
	B [177 214 253]	b [115]
	C [30 145 108]	c [107]
C	A [221 236 220]	a [82]
	B [10 8 8]	b [82] and [90]
D	A [132 48 18]	a [148]
	B [181 213 229]	b [82]
	C [43 154 104]	c [82]

## 5.9 Conclusion

The main contribution of this chapter is in the field of metrology and machine vision by developing a method of using low-cost Lytro cameras in metric depth measurement. The factors which may affect the measurement of the Lytro cameras such as the lighting conditions and object surface have been investigated. A new way of measuring SI unit data from LF cameras which produce a greyscale varying depth map called response curve has been introduced. Response curves illustrate the relationship between greyscale depth data from the camera with SI unit distances. Also, the pixel resolution of the Lytro family cameras was calculated which enables the measurement of any

scene with 3D data in absolute metric units. Furthermore, the depth sensing range for all three Lytro cameras has been defined which range from 50 mm to 270 mm for the Lytro - I generation cameras and 30 mm to 680 mm for the Illum camera. This data will help users to prejudge the kind of application their cameras can be used to generate absolute 3D data. Along with the above results, some of the noticeable results in this work are:

1. The accuracy and repeatability achieved were +10.0 mm to -20.0 mm, and typically 0.5 mm respectively in Z coordinate (depth) since response curve was generated at 5 mm intervals and hence accuracy is closely related to intervals of the response curve.
2. For the lateral X and Y coordinates measurement, the accuracy was +1.5 mm within the active zone of cameras
3. The uniform illumination experiment demonstrates the light intensity dependency of the Lytro camera to generate a depth map along with some other variable that triggers depth results to be different than expected (active zone size changes).

The results of this work are promising to help enable the use of LF camera in metrological applications with limited freedom with respect to range and accuracy. The user must invest time in generating response curves and pixel resolution to use the Lytro cameras to generate absolute depth in a specific environment. In this work, only the RGB data and greyscale depth map were used to generate results, but the LDS generates other important results that can be used to generate useful information for engineering applications, such as perspective images. Chapter 6 deals with information available using the LDS generated perspective views and the procedures to generate absolute depth maps from data recorded using the Lytro cameras.



# 6



ABSOLUTE DEPTH  
USING  
STEREO-VIEW DATA FROM LYTRO  
CAMERAS

## *Overview*

Lytro family of the camera are low-cost cameras capable of capturing light radiance as exhibited by expensive LF cameras. Apart from being cheap, the Lytro cameras are designed with interesting hardware features such as touch to focus, changing focal length, field-of-view and many more. In addition, the Lytro cameras are provided with free software (Lytro Desktop tool) to gain access to the LF captured by generating a relative depth map, sub-aperture images, metadata file and in few other forms.

The results generated by Lytro desktop software was explored in Chapter 5, by making use of relative depth maps. Using the response curve, it was made possible to transform relative depth maps into an absolute scale. But, no great attention has noticed by the engineering community in exploring other results provided by Lytro software. Many research teams have claimed to use laboratory built LF cameras for generating various results. Since this work mainly targets commercially available low-cost LF cameras (Lytro), we make use of other results from Lytro software in this Chapter.

Furthermore, in Chapter 5, it was noticed that Lytro-I generation cameras to have limited work volume compared to the Illum cameras. Since relative depth results were used in Chapter 5, evidence-based claims were not possible. Hence in this Chapter, the sub-aperture images are used to determine the performance of the Lytro family of cameras. The sub-aperture images are transformed results of actual LF radiance; many interesting observations are made in this Chapter. The relation between base disparity value and touch-to-focus feature act as the key results to obtain absolute depth data from the Lytro captured images. Also, the work volume relationship with the focus plane is discussed in this chapter.

In this Chapter, two methods are suggested that helps to generate absolute data in SI unit by targeting some of the pros and cons exhibited by the Lytro family of cameras. The drawbacks mentioned point towards the need for an LF calibration method and user-controlled LF software that lead to Chapter 7 and 8.



## **6 Absolute Depth Using Stereo View Data from Lytro Cameras**

### **6.1. Introduction**

Over the last few years, there has been great interest in developing and understanding LF technology (Plenoptic) for different applications such as photography and computer vision. The photography community generally looks for enhancements in camera technology (LF) that bring flexibility in capturing the desired scene with adjustable camera features such as focal length, aperture and zoom. However, the engineering community, especially computer vision groups, try to make use of additional information recorded by LF cameras. There have been few applications of using LF cameras in metrology applications and very little research has been published related to low-cost LF cameras (Lytro) regarding their advantages and drawbacks in capturing scene data in absolute scale measurements.

LF cameras capture light radiance along with the direction using an MLA. The additional optical element in an LF camera (MLA) introduces a problem with respect to the baseline distance to calculate depth. The baseline can be defined as some fixed (known measurement) distance between two cameras or camera views, that help to convert the disparity map into a depth map in absolute units. In the case of MLA based LF cameras, the pitch value of the lenslet units are measured in micrometres (13.89  $\mu\text{m}$  for the Lytro-I generation camera) when compared to conventional cameras that have a baseline in the range of few millimetres to metres (depending upon application [94–96]). In stereo-vision techniques [102], the baseline distance between two cameras is physically measured and incorporated for calculating disparity in absolute units under constant focal length condition. Physically measuring baseline distance between camera views in MLA based LF cameras is not feasible, since the MLA is not accessible once the MLA is assembled with the photosensor. Also, any baseline calculation made before the MLA was assembled with the photosensor may vary depending upon the distance between the MLA and the photosensor. Hence absolute disparity calculation using an LF camera is challenging.

Existing research results use different methods such as: generating focal stacks [103] from 4D LF data, defocus/focus measure [104], combine defocus cue and correspondence data [14, 100] to generate a relative depth map (not in absolute units). A hybrid camera system was designed by combining a DSLR and a Lytro camera to generate images with high dynamic range and resolution, along with a depth map using the DSLR image and central view of Lytro cameras [49], which resembles stereo-vision technique. Furthermore, most of the published results regarding techniques to generate a depth map using LF techniques tend to use laboratory build LF cameras with fixed focal length. Only a few research groups have considered commercially available low-cost LF cameras (Lytro) to obtain metric depth map (see Chapter 5) [95]. For instance, the depth map generated by the Lytro family of cameras was calibrated up to 1,000 mm in Chapter 5, but results were restricted to specific colour targets and working environments (Illumination), while no specific use of baseline values was made. A previous prediction of baseline positions based on the parameters of a standard virtual LF camera setup provides the proof of using virtual distance between the camera views (perspective light rays) as standard baseline and results were measured with standard ray tracing software with less than 0.2% error [106], but no practical implementation was demonstrated.

This chapter plays the role of bridging the gap between commercially available Lytro cameras (Lytro-I generation and Illum) and new techniques to generate absolute scale depth maps with high user flexibility in terms of selecting the work environment and lighting conditions. In addition, this chapter considers using existing camera models along with software packages to enhance the end results along with reducing the total time required to calibrate the camera to suit the work environment. Furthermore, additional factors affecting the Lytro cameras to provide accurate depth results are investigated and alternative techniques to overcome these factors to use the Lytro cameras efficiently for metrological applications are suggested.

Some of the notable advantages of using Lytro cameras are:

1. The Lytro cameras are among the few commercially available LF cameras capable of recording 4D information in a single operation. Apart from availability, these cameras are affordable in the range of £ 150 to £ 1,000. To access the features of LF technology, the Lytro company also provide a software

package at no extra cost, known as Lytro Desktop software (LDS). This helps the user to save the time invested in developing an LF cameras model using MLA and related software package.

2. The Lytro cameras are designed to perform as DSLR cameras by providing many features to the users such as digital zoom, changeable focal length, shutter control and touch to focus (TTF). This has been made available by a well-designed compound optical system supported by mechanical design. Most of the mechanical and optical parameters used in the design of the Lytro cameras are available (metadata of Lytro image consist of camera parameters used while capturing images and few physical camera parameters are available from Lytro Beta package) thereby reducing the human hours in calculating these parameters for any application.
3. The focal length and zoom features in the Lytro cameras are flexible and can be changed to match the user requirement. This feature enhances the chances of using the Lytro camera in a wide variety of applications since there are only a few LF cameras providing these features at the low-cost margin.
4. The data obtained by the Lytro software is of high quality (in terms of RGB and resolution) since these images undergo spatial correction of pixels in 2D coordinate plane, i.e. in an image (specific details of spatial correction of pixels are unknown since the Lytro company has not disclosed this information), colour correction and enhancement to match the captured scene.

In Figure 6.1, it can be noticed that the image set generated by the Lytro software is of good quality compared to the manually generated image set from the same raw file captured by the Lytro cameras. The greyscale values of a scan line are compared in Figure 6.2, where the intensity values are compared for an 8-bit image (0-255). For stereo matching applications, a high dynamic range image with a good contrast (Figure 6.1 top) would result in better results compared with dull and low contrast images (Figure 6.1 bottom) which reduce the accuracy of matching algorithms.

In Figure 6.2, the blue line represents the intensity values for the LDS generated image and has a sharp distinction between white-black regions (background, goggles) which range between greyscale values of 20 and 180. While compared with the red line representing the intensity values of the manually generated

image ranging between 90 and 130 greyscale values. Therefore, in this example manually generated images required additional time to enhance the image data which is available with the Lytro software at no additional charges.

Figure 6.3 represents the disparity estimation results obtained by considering two perspective images generate by LDS and manual method. Here, the word disparity refers to the distance between two corresponding pixels/features in the left and right image of a stereo pair. The disparity estimation provides consistent values for LDS data, while the discontinuous result for manually generated images. It can be noticed in Figure 6.3 that disparity values for both results vary and this may be due to uncalibrated manual data compared with calibrated LDS results. Due to high contrast and distinctive pixel values, images from LDS provide better matching results when compared with manually generated images of Figure 6.1.

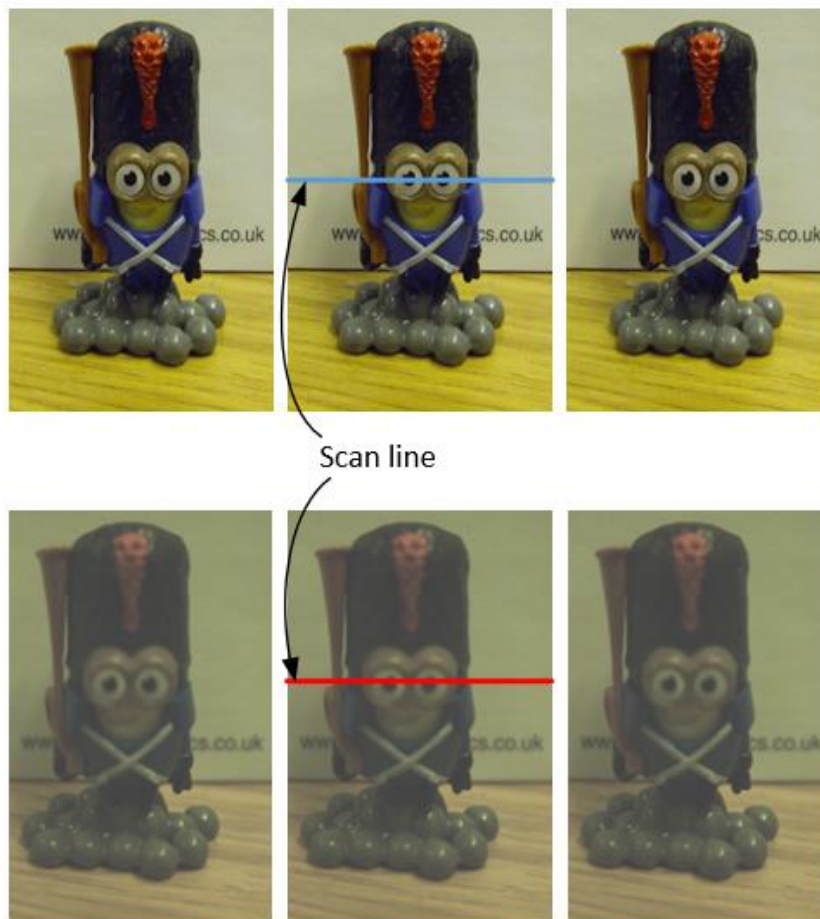


Figure 6.1 A set of three perspective images generated using the Lytro desktop software (top) compared the same perspective images generated manually (bottom)

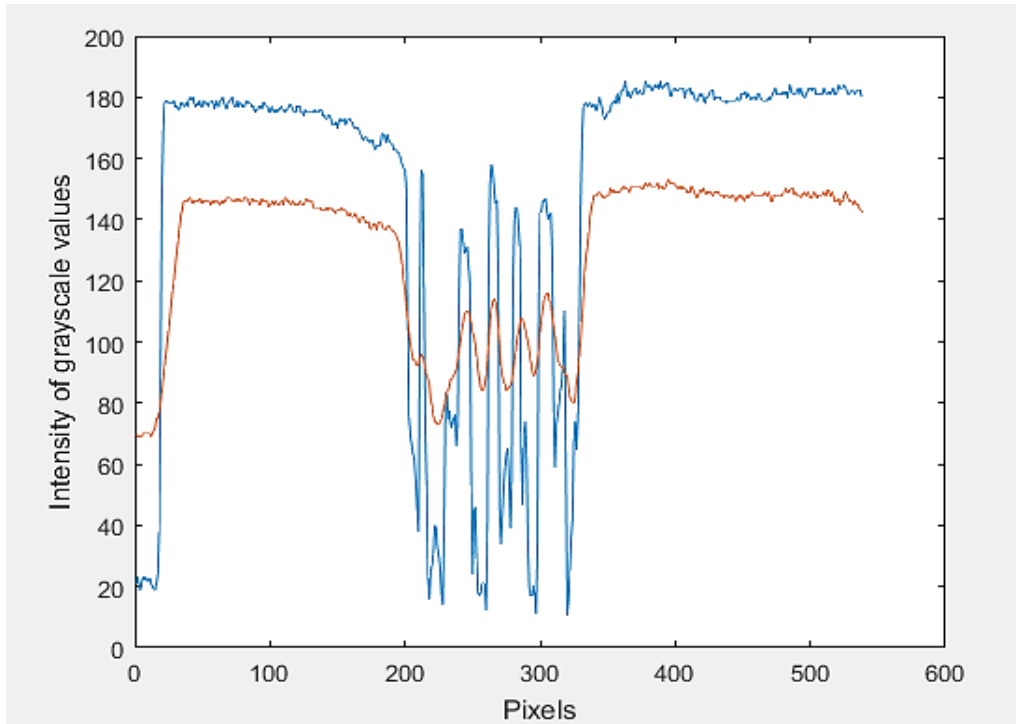


Figure 6.2 Representation of intensity values of the scan line shown in Figure 6.1

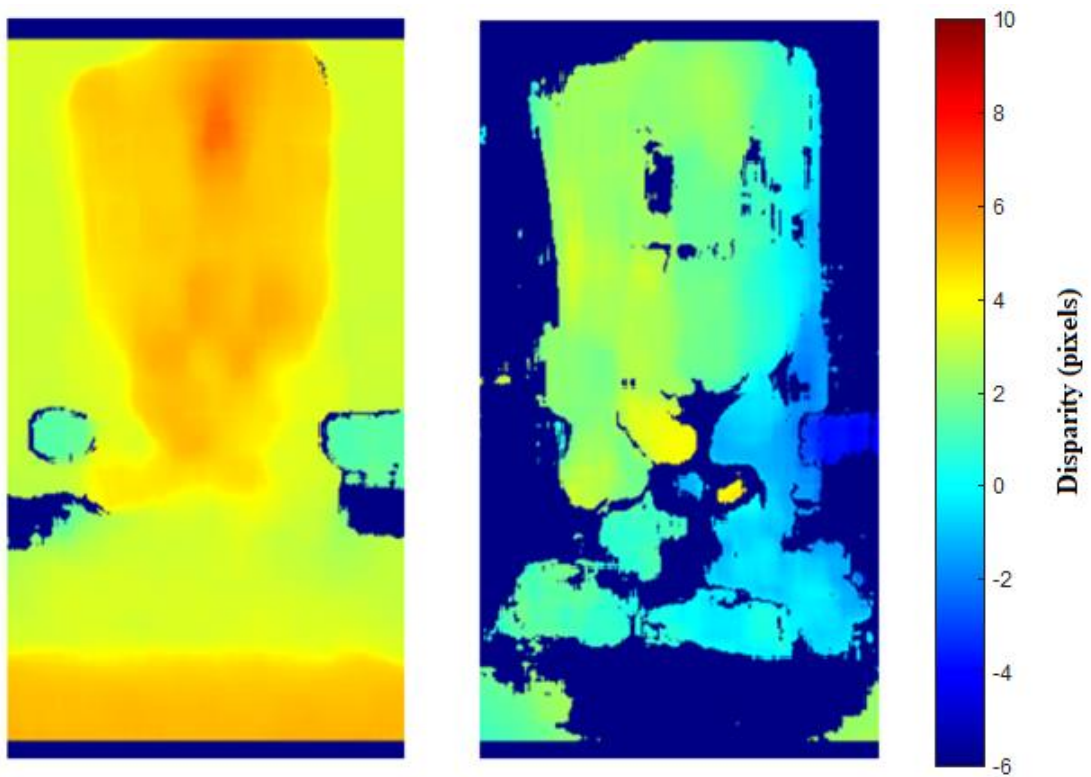


Figure 6.3 Disparity estimation of the Lytro images (left) and manually generated views (right).

## 6.2. Theory of perspective views

The perspective views are also known as sub-aperture images that represent the 4D LF data captured using MLA based cameras as 2D camera views (see chapter 2). The perspective views are generated by selecting the light rays of a direction from all individual MLA lenslet and arranging according to the MLA locations. These perspective views can be considered as images from an individual array of camera units capturing the same scene with slightly different orientations that depend on the shape of the MLA used [66, 102, 103]. The number of perspective views generated using an LF camera depends solely on the micro image pixel count (see Chapter 3). These perspective images can be used with the central view to generate a stereo-vision system, thereby calculating the disparity of the scene.

In the case of the Lytro cameras, the perspective view generated by the LDS are all-in-focus images. The LDS pre-processes the captured LF data by generating multiple focal stacks and calculates the relative depth map, thereby generating information that indicates the best-focused  $Z$ -plane for every pixel in the image. Thus, the perspective view generated by the LDS can be used to generate a relative depth map but not an absolute depth map (e.g. in SI Units).

Apart from generating all-in-focus results, the Lytro cameras inherit some basic features from conventional cameras. When an image is taken using a Lytro camera the lens system is adjusted to bring the subject into the high contrast between the foreground and background. In this situation, the raw LF image captured by the camera demonstrates sharp details of the subject (focused region) and blurry details of foreground and background. This process is similar to conventional cameras, but due to the presence of the MLA helps in recording the direction of light along with the positional information thereby transforming the out-of-focus regions into in-focus [109].

This mechanism is illustrated in Figure 6.4, where two-point objects  $o$  and  $o'$  (subjects) are captured using a Lytro cameras with the TTF feature enabled. The subjects  $o$  and  $o'$  are placed at a fixed distance of  $u$  and  $u + \Delta u$  respectively. The compound lens system (CLS) consists of two stepper motors to adjust the zoom and focus lens respectively to adjust user defined field-of-view and focus range. For ease of visibility, two perspective views (shown in green and red) along with a central view

(shown in black) are considered. In the partial scenario, the perspective views are very close to the central view, while the distance between MLA and the photosensor is short but in Figure 6.4 both of these are exaggerated for better visibility. It is noticeable that, the perspective views are being focused at a single point behind the lens assembly, at a distance equivalent to the focal length of the CLS thereby creating a distance of  $d_1$  and  $d_2$  between the central view and corresponding perspective views [106].

When the subject distance from the camera plane is increased to  $\Delta u$ , the internal motor steps move the CLS to generate a focused image of the subject on the image plane. In this situation, the focal length of the CLS changes to  $f + \Delta f$  (this is shown in Figure 6.4 by changing the distance between internal lenses), since the MLA and sensor assembly (just the sensor assembly, in the case of conventional cameras) is fixed. This results in a distance of  $d_1$  and  $d_2$  between the central view and the perspective views at a distance equivalent to the focal length of the CLS. Due to this mechanism, the perspective views are further channelised (i.e. perspective views move further apart, less chances of interaction between two perspective view light rays) to be registered on the photosensor pixels with the help of the MLA where the user selected plane is at focus compared to other  $Z$  planes from the camera.

In comparison with the conventional cameras, the photosensor assembly at the image plane (or few millimetres further away from the image plane) force the perspective views to merge, resulting in a single focused  $Z$  plane. Thus, images from conventional cameras lack perspective view details that are consequently represented as a blur. Although it should be noted that some autofocus cameras make use of two perspective views to generate focus readings using a split prism mechanism [110].

Since there exists a constant distance  $d_1$  and  $d_2$  between central and perspective views (shown by enlarged views, see Figure 6.4) which can be used as a reference to describe the disparity map in absolute units. The constant distance is known as the base disparity ( $b_d$ ) that remains constant for the  $Z$  plane selected by the CLS and TTF. The base disparity value is expressed in pixel units, since it is the measure of a number of pixels between the perspective views and the central view. This value can be used as a reference to calibrate the disparity map obtained using any two stereo-pair images of the Lytro cameras (a perspective view + central view).

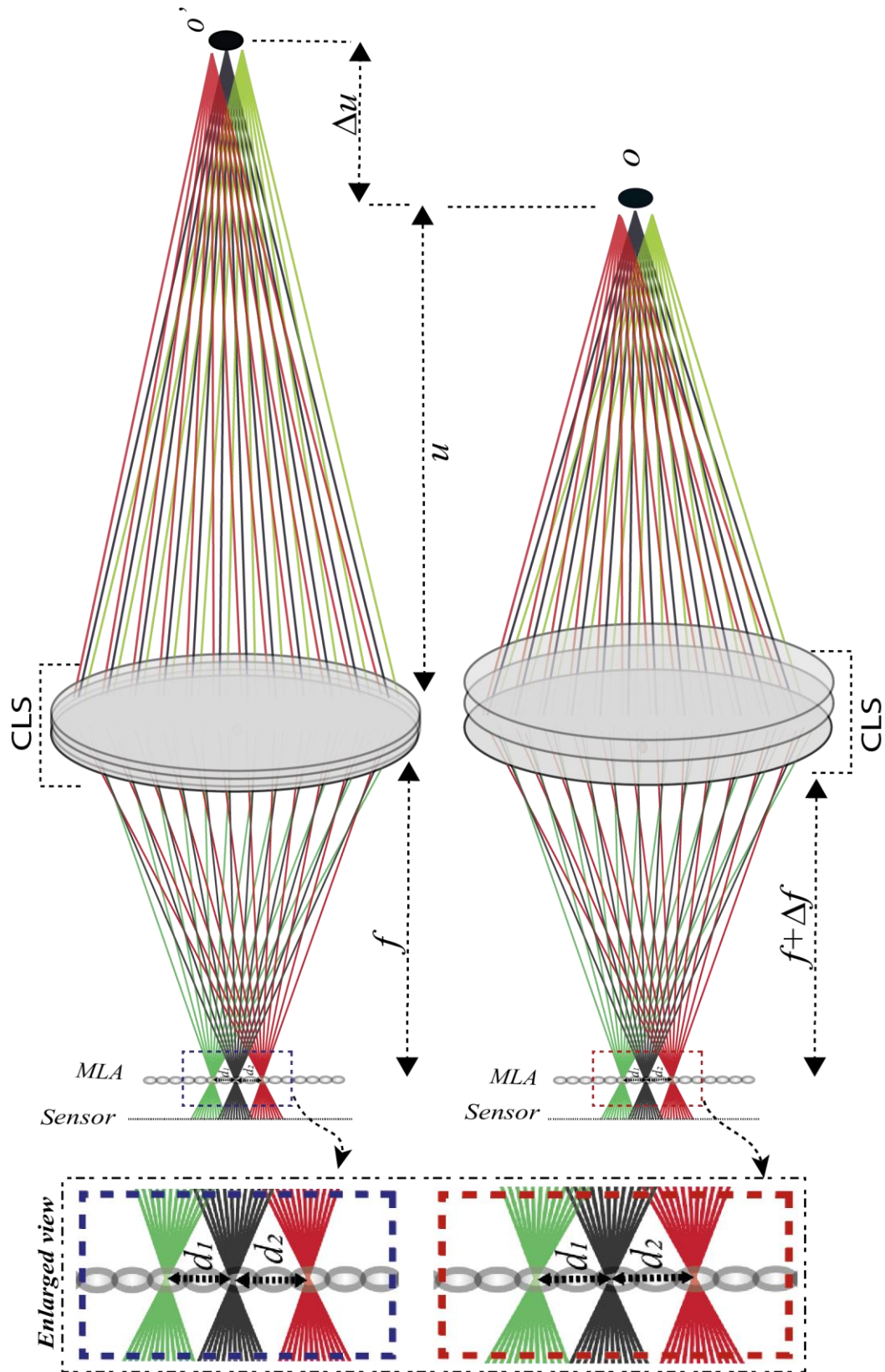


Figure 6.4 Illustrating the internal compound lens adjustments to capture point object (subject) placed in front of the camera resulting in a constant base disparity ( $d_1, d_2$ ) in the perspective views regardless of the subject distance



If the user selected  $Z$  plane distance is known, then the base disparity is converted into the absolute scale and hence the entire scene captured using the Lytro cameras can be converted into absolute units. The user selected  $Z$  plane distance can be generated by evaluating the stepper motor values of the zoom lens ( $ZS$ ) and focal lens ( $FS$ ) of the Lytro family of cameras (see section 6.5).

### 6.3. Order of views

The LDS generates 7 images ( $p_v$ ) representing the perspective view of the scene captured by the Lytro family of cameras. Regardless of the camera used (Lytro-I generation or Illum), the Lytro software produces 7 different views of the scene that can be used for generating an absolute depth map. Since the Lytro Company has not disclosed the work-flow of the LDS engine nor is any information available online, hence the LDS has been considered as a black box that generates output for some given LF input.

To make use of these perspective images, the arrangement of these images with respect to the MLA configuration needs to be calculated. From chapters 3 to 5, the hexagonal structure of the MLA used in the Lytro family of cameras is known. With this knowledge, the perspective images generated by the LDS were expected to be in the hexagonal pattern. The perspective views ( $p_v$ ) generated by the LDS were provided with image file names, with a specific name starting from 00 to 06 as shown in Figure 6.5. However, the LDS did not provide any information to consider the order of the file names corresponding to the MLA structure, i.e. for example, no evidence was provided to consider the image file 01 and 02 as stereo-pairs (left and right perspective views).

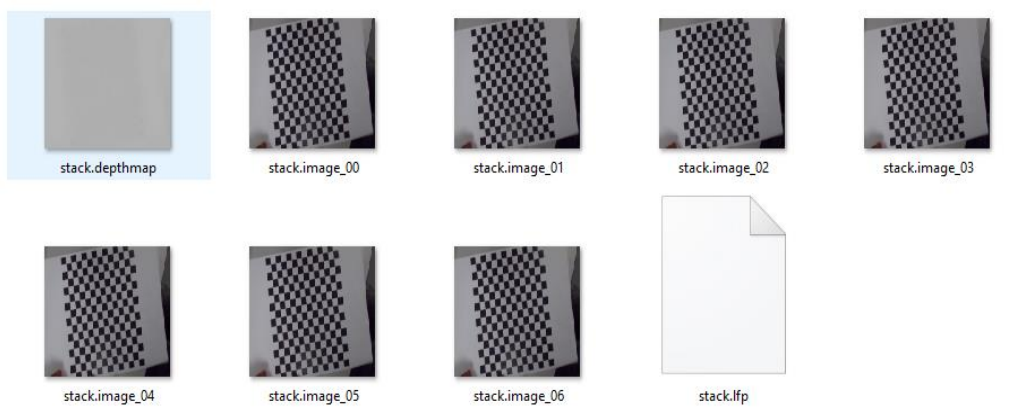


Figure 6.5 Results generated by Lytro Desktop software under 'Editable Living Pictures Mode'

The stereo-pair information defines a rigid relationship between two images that remains unchanged for a given camera system. If the stereo-pair relation of camera views is known, then the distance between these views (baseline distance in conventional cameras [99]) can be used as a reference to generate useful data (example, absolute depth map). Further, if the stereo-pair relation is an only known parameter, then by comparing images (disparity estimation) a value equivalent to baseline can be generated in terms of pixels. Hence to generate evidence-based information of the stereo-pairs, a white-black checkerboard was captured using the Lytro family of cameras. The Lytro cameras (LC1 and Illum) were placed perpendicular to the checkboard so that further calculation of comparing checkerboard corner location within the perspective views was made straightforward. This is because the perspective images are the transformed LF data with respect to the direction of light rays. The experimental setup is shown in Figure 6.6, where  $CB$  represents the checkerboard under consideration. The Illum camera was replaced with the Lytro-I generation whilst the rest of the experimental setup remained unchanged. Let  $o$  be the optical centre with respect to the Lytro camera,  $d_z$  be the variable distance between  $CB$  and the exit pupil of the Lytro family of cameras with  $d_c$  as the constant distance to the photosensor from the exit pupil.

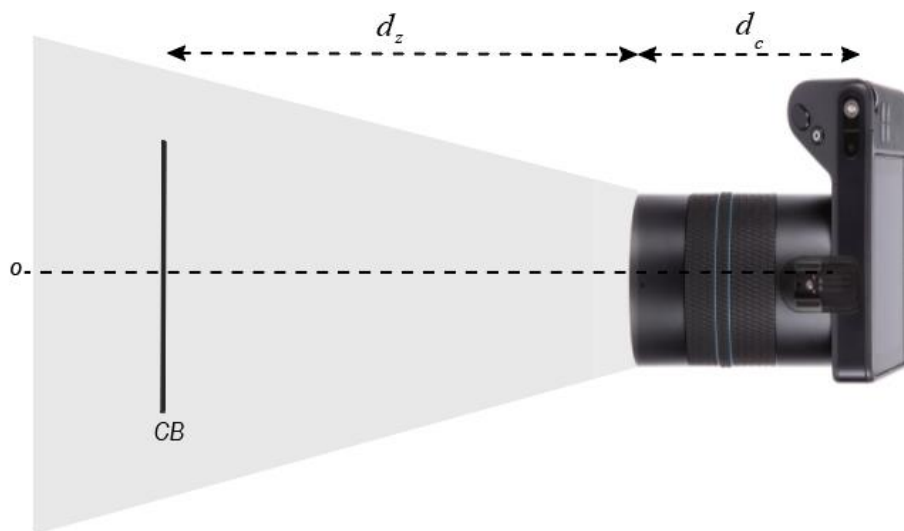


Figure 6.6 Experimental setup to calculate the sequential order of the perspective views generated by the Lytro Desktop Software

The perspective images of the checkerboard were generated using the LDS and corner points were calculated with respect to the image coordinate. The 00-image file was considered as the central view of the MLA structure and corner points from the rest of the perspective images are marked on the 00-image file (overlapped), as shown in Figure 6.7. The perspective views shift 180 degrees in the results of both the Lytro family of cameras, i.e. the location of the 05-perspective image is shifted 180 degrees (see Figure 6.7 top row) for Illum camera and vice versa (see Figure 6.7 bottom row) for Lytro-I generation camera. This behaviour suggests that there was no consistency used by the LDS in allotting the numbers for perspective view images, hence causing problems when selecting images for stereo-pair disparity calculation. It can also be noticed that the perspective representation of the corner data in Figure 6.7 results in a hexagonal shape which provides further evidence of the hexagonal MLA structure of the Lytro cameras.

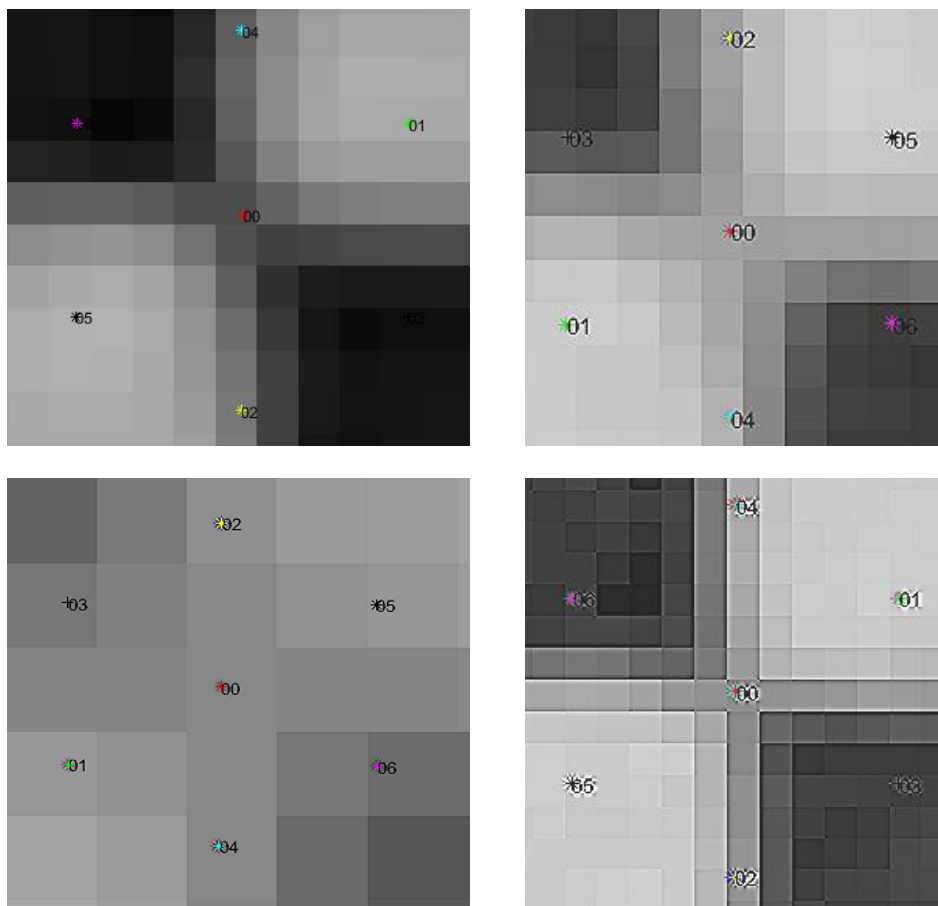


Figure 6.7 Overlapped corner points of perspective views with the 00-image file as a central view for Illum camera (top row) and Lytro-I generation camera (bottom row). Numbers 00 to 06 represent the perspective views (perspective view image files) of the Lytro family of cameras considering 00 as the central view.

#### 6.4. Importance of touch to focus

The TTF feature is an inbuilt element of the Lytro family of cameras that can be used to select a plane in the field-of-view of the camera. With the help of this feature, objects of interest (OOI) are brought into sharp focus on the Lytro camera display. The Lytro cameras can generate all-in-focus images after processing the LF data captured by the cameras, but the TTF feature will change the internal focal lens and zoom lens motors steps to bring the OOI into high sharpness, i.e. this feature will ensure the light rays from the selected plane converge on the photosensor.

A group of 20 checkerboard images were captured using the Lytro family of cameras (Lytro-I and Illum) and the perspective images were exported using the LDS. The checkerboard corner was detected in all perspective images and plotted with one of the perspective views as a reference to generate the sequential order information (similar to the procedure used in Figure 6.7 ), but the checkerboard was manually selected using the TTF by the user before capturing data with the Lytro family of cameras. Figure 6.8 illustrates the outcome of the experimentation and it can be observed that irrespective of the perspective view considered as a reference image (identified on top of images, also marked with a dotted line and located at image coordinate (0,0)), the sequential order remains constant. The resulting hexagonal perspective view with repeating sequential order provides the evidence for stereo-pair disparity calculations defined by using the TTF.

The resulting sequential perspective images are shown in Figure 6.8 (bottom right). The 00 view of the Lytro camera is the central view and views 04 and 02 representing the top and bottom pair respectively. While, views 01 and 03, and 06 and 05 represent the right and left pair respectively. Along with stereo-view information, it can also be observed that distance between central and side view generates a small distance marked as  $x$  and  $y$  in Figure 6.8 (bottom right) which represent the baseline in pixels and this information plays an important role in generating absolute depth maps from the Lytro camera perspective data.

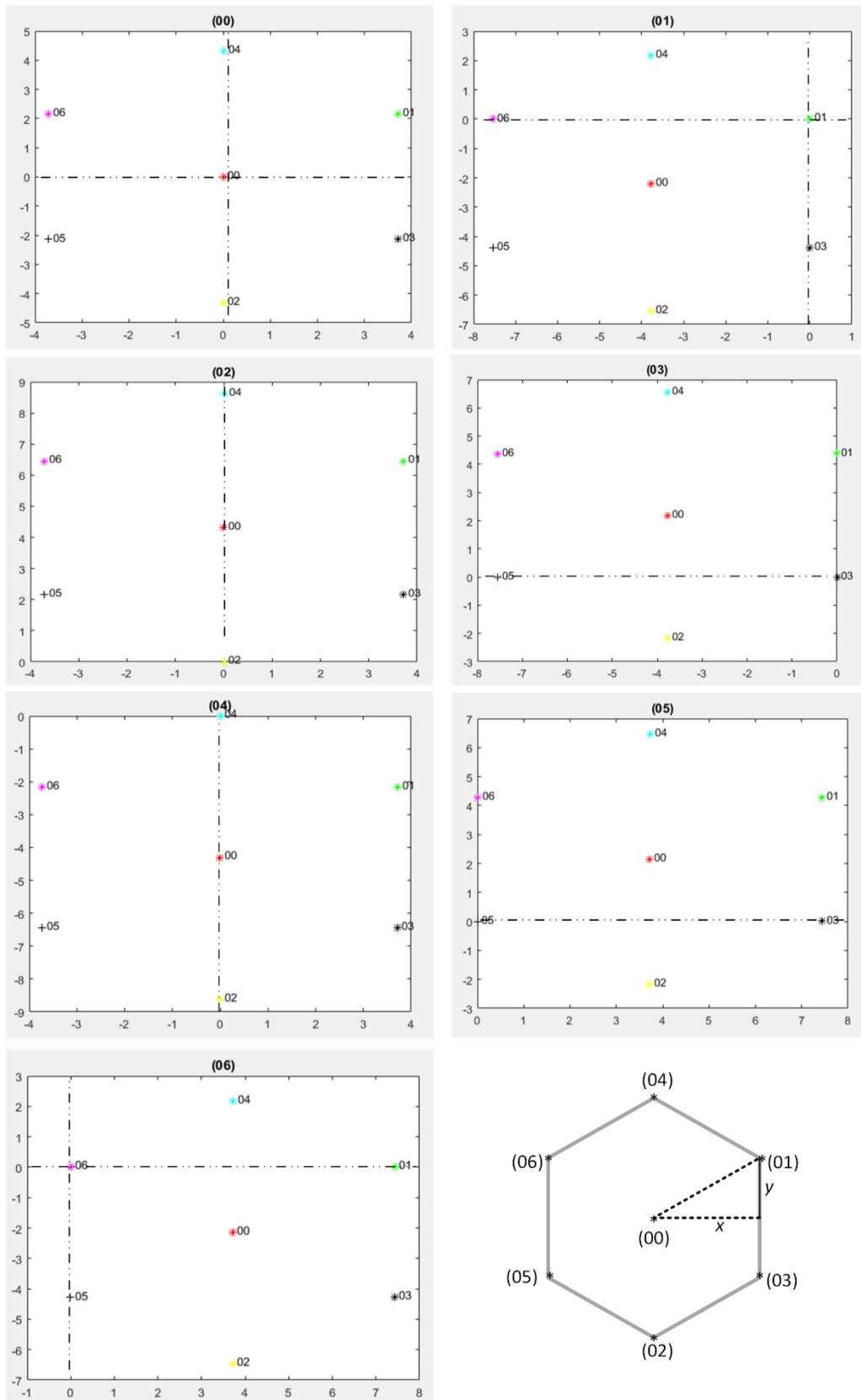


Figure 6.8 Representation of the sequential order of perspective images with respect to reference image marked at location (0,0) and resulting perspective orientation chart (sequential order) (bottom right)

Comparing a central view (00) with any one of the perspective view images (01 to 06), they are shifted by a few pixels which are equal to the pixel distance of their respective light direction considered while transforming LF data into a perspective image. For example, if the 00 images are generated by considering a central light ray of all individual MLA units and the 01 images are generated by considering a 5th light ray of all MLA units, then the pixel shift on the perspective images is equal to 5 pixels when both images are overlapped. This forms the basic concept of finding the baseline in terms of pixels (base disparity) for stereo-pair images.

### 6.5. Base disparity between perspective views

Using a pair of conventional cameras as a stereo-pair assembly is a well-known technique to calculate depth data of a given scene [94, 95]. Important key terms used in this method are focal length ( $f$ ), baseline ( $b$ ) and disparity ( $d$ ), resulting in  $Z$ -axis depth calculation as given by Equation 6.1. Under the condition of calculating the baseline of stereo-pairs with known object distance ( $Z$ ), focal length and disparity, it would be possible to determine the base line by transforming Equation 6.1 into Equation 6.2. This is because conventional cameras use a fixed focal length to produce an image as shown in Figure 6.9.

$$Z = \frac{f b}{d} \quad 6.1$$

$$b = \frac{z d}{f} \quad 6.2$$

Let  $f$  be the focal length of the lens used to image an object ( $CB$ ) on the photosensor placed at a distance of  $v$  from the lens as shown in Figure 6.9 . Under this condition, the distance to generate a focused image of the object will be at the image plane  $i$ , at a distance of  $u$  from the lens. If the object is placed at distance  $u + \Delta u$ , the resulting image is blurred, and conventional cameras do not have any additional data to enhance the blurred image. Furthermore, the blurred data can be used to calculate the disparity map and hence the depth data due to constant focal length. But, the depth recovery is limited and less accurate due to the blur effect.

In contrast, using the MLA helps LF cameras to record additional directional information on the 2D photosensor along with the intensity data. Using the directional information, the virtual focal lengths ( $f + \Delta f$ ) are generated with the help of MLA to obtain sharp pixel data, that were lost due to the objects locations ( $u + \Delta u$ ). Hence, the

output image is not related to a fixed focal length used by LF camera while capturing the image. A sharp image is obtained as output (see Figure 6.9 (bottom)) for all object distances of  $u + \Delta u$ . Table 6-1 shows the results of using Equation 6.2 to calculate the baseline, where constant focal length was used to calculate the baseline for four targets at varying distances from the camera. The disparity involved in the perspective view images generated by the LDS was calculated using Equation 6.3, where  $LI$  and  $RI$  are the pixel coordinates of perspective views (images) under consideration, while subscript  $h$  and  $w$  represent the disparity direction with respect to image origin (0,0). The resulting baseline values are not constant, indicating the changing focal length effect in the perspective views generated by the LDS.

It can be noticed from Table 6-1 that generating the baseline similar to the conventional stereo-vision method is not possible with LF images even under constant focal length. But as explained in section 6.2 and the virtual simulation of LF cameras in [106], suggests that Lytro cameras possess a constant disparity between perspective views. Hence to demonstrate the fixed disparity an experiment with three checkerboards was performed as explained in Figure 6.10 and Figure 6.11.

$$\begin{aligned}
 \text{height} &= LI_h - RI_h \\
 \text{width} &= LI_w - RI_w \\
 \text{disparity} &= \sqrt{\text{height}^2 + \text{width}^2}
 \end{aligned}
 \tag{6.3}$$

Table 6-1 Representation of values used for calculating baseline using Eq 6.2 for Lytro family of cameras with perspective view 00 and 01 under consideration

Focal length (mm)	Disparity (pixels)	Object distance (mm)	Baseline (pixels)
9.4	2.89	0.16	49.22
9.4	4.12	0.21	92.19
9.4	4.39	0.27	109.35
9.4	5.16	0.36	171.51
2.4	37.22	0.36	1080.72
12.4	16.3	0.26	343.68
12.4	0.82	0.16	10.66
12.4	6.91	0.48	269.24

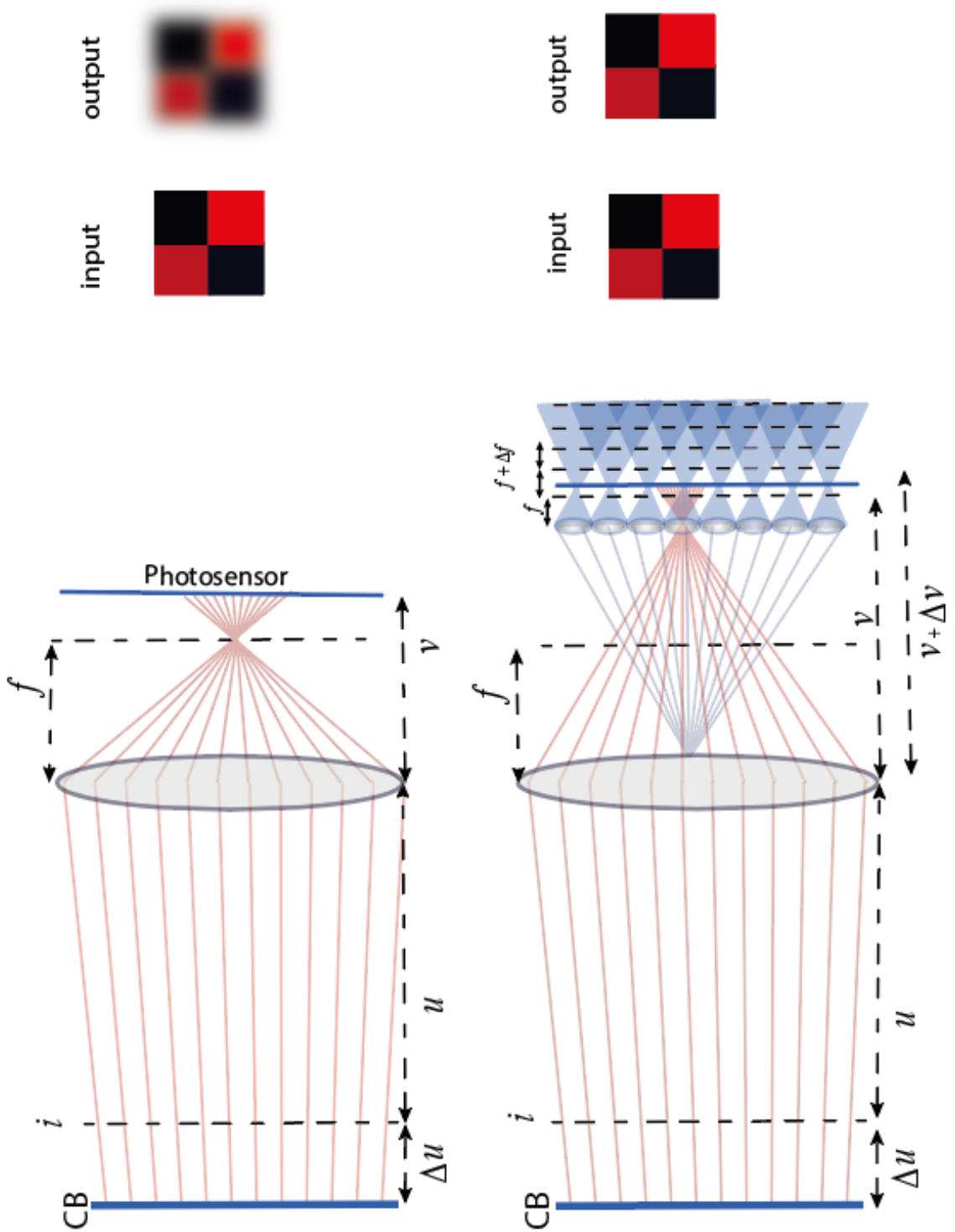


Figure 6.9 Image formation in a conventional fixed focal length camera (right) and in a microlens array based LF camera (left) (figure not to scale)



The perspective images generated by the LDS are the transformed 4D LF images into 2D camera views, hence light interaction inside the camera can be simulated as if the Lytro cameras were replaced with multiple 2D cameras system. In Figure 6.10, three checkerboard planes were considered, CB1 to CB3, while CM1 to CM3 are the perspective views of the Lytro (equivalent to the side-by-side setup of three 2D cameras) represented by their light rays interacting with the sensor. In Figure 6.10, the photosensor is overlapped with the checkerboard planes to highlight the converging (bright-thick lines) and diverging (faded-dotted lines) light rays with respect to the planes selected by the user. When the user selects the plane CB1 using the TTF, the CLS is adjusted automatically to get the CB1 surface at the sharper focus. During this process, the light rays converge at a point to generate a minimum distance between light rays of the neighbouring views, represented by  $b_d$ . If the value of  $b_d$  remains constant with change in user-selected plane, this can be used as a base reference to generate depth map in absolute scale. The faded lines beyond the sensor represent the region where LF techniques are used to recover the scene details. By calculating the disparity at CB2 and CB3 surfaces, a relation between rates of change in the disparity values at various distances from the focused plane can be generated. The process is repeated when the user selects the CB2 or CB3 planes to be focused by the CLS and the minimum distance between perspective views is calculated, known as the base disparity( $b_d$ ), along with the rate of change in disparity related to distance.

The Lytro family of cameras (Lytro-I and Illum) were arranged in a fashion such that optical axis of cameras was perpendicular to target objects as shown in Figure 6.11, to calculate the base disparity values as illustrated in Figure 6.10 with the TTF feature activated. While capturing images, care was taken to manually select the target by tapping on the live preview display (TTF) of the Lytro cameras. Furthermore, the experimental setup is shown in Figure 6.11 allowed for the cross verification of the matching accuracy between camera focused plane and user selected plane, since no direct method exists currently to illustrate this effect. Also, the matching accuracy of the Lytro family of cameras describes the CLS efficiency to distinguish between two target planes with high contrast involved between them.

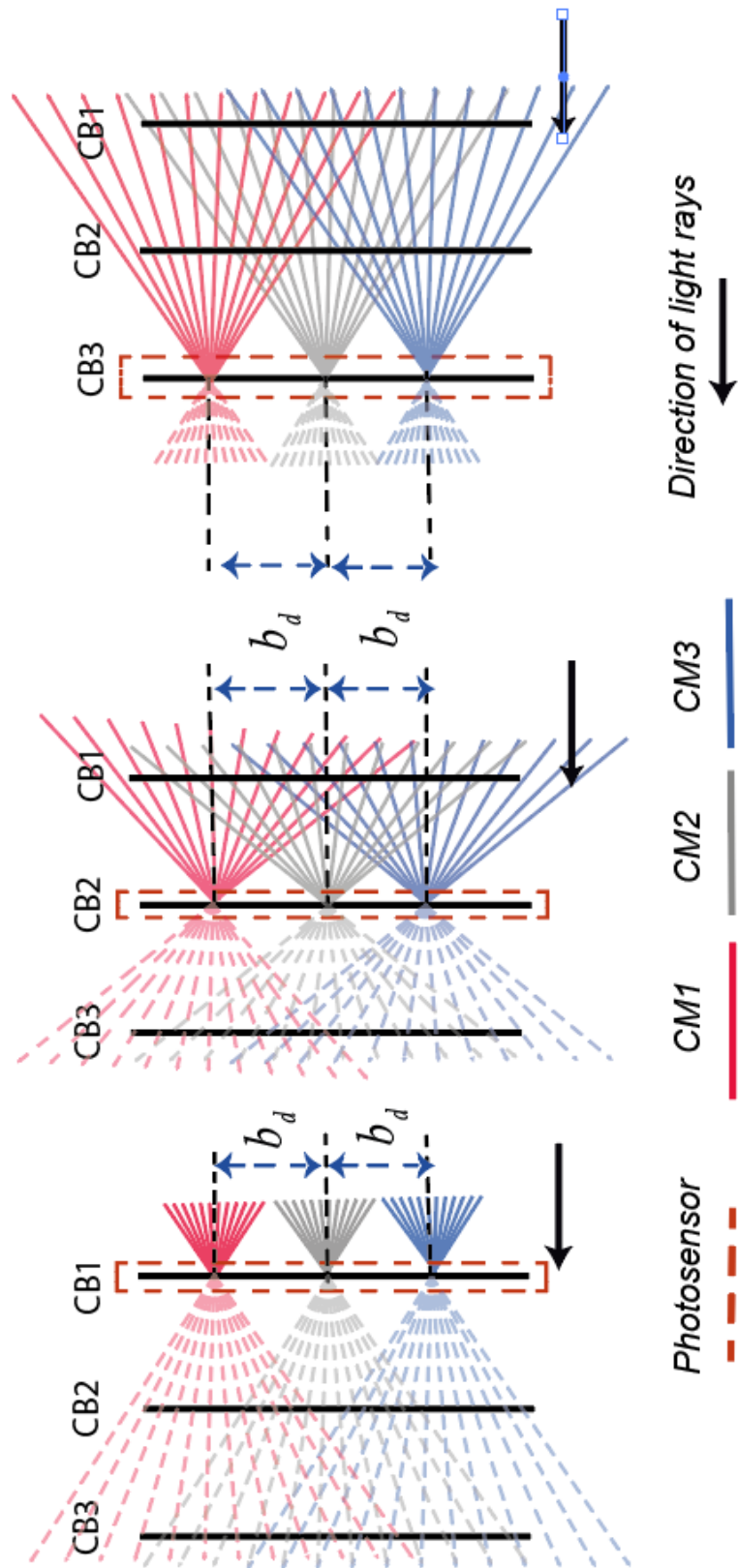


Figure 6.10 Interaction of light rays in a conventional multi-camera system with the photosensor under user-selected plane at the sharp focus

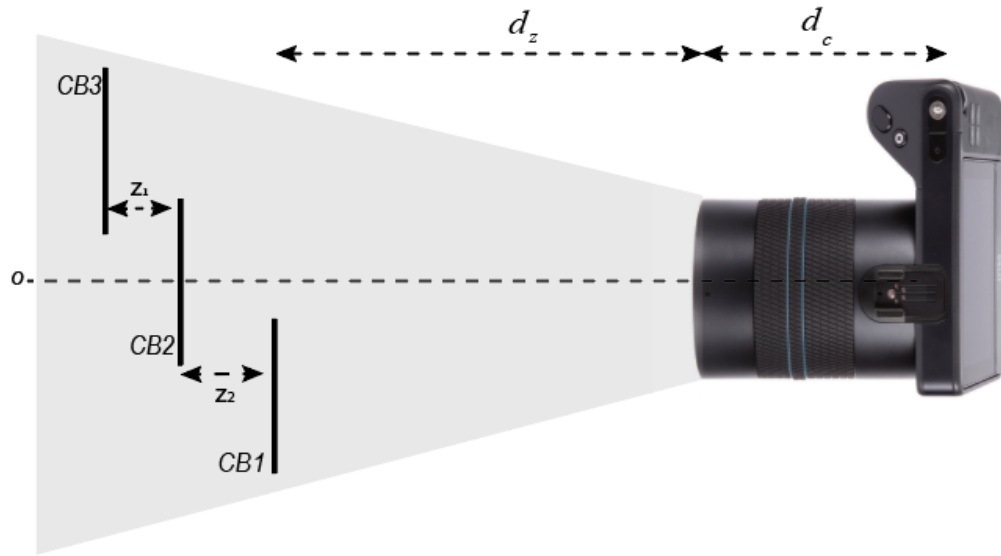


Figure 6.11 Experimental setup to illustrate the importance of ‘Touch to Focus’ feature and gather the information of light propagation before and after the TTF line

In Figure 6.11,  $Z_1$  and  $Z_2$  are the distances between checkerboards CB1 and CB2, and CB2 and CB3 respectively, while  $d_z$  and  $d_c$  are as explained in Section 6.3. The experimental setup remains the same while cameras Lytro-I and Illum were swapped for collecting images to calculate base disparity values.

Using the TTF feature, checkerboards CB1 to CB3 were selected in sequential order and sets of five images were captured. CB3 was placed at the far end distance used by this experiment, 800 mm, while the CB1 was placed at 200 mm, the minimum distance between checkerboards and cameras, due to experimental set-up constraints. CB2 was placed at an incrementing step of 50 mm from CB1 and corresponding sets of five images were captured for each step. The obtained image set was subsequently processed using the LDS to generate the perspective views.

To obtain the disparity between perspective images (image file named 01 to 06) with reference to the central view (image file named 00), the checkerboard detection method used in [106, 107] was employed to generate the checkerboard data of different grid sizes. The grid size of checkerboard CB1, CB2 and CB3 were 19.1 mm, 25 mm and 35 mm respectively. The variation in the size of the checkerboard was to accommodate maximum grids when checkerboards were close to the camera and to ease the detection of corners when the checkerboards were far away from the cameras. The pair of images under consideration, for example, 00 and 01 images were processed to generate the

checkerboard corner data  $P_0$  and  $P_1$ . Since the perspective views were shifted by a few pixels compared to the central view, some of the corner points present in  $P_0$  did not match with  $P_1$  due to the occlusion effect and noise (e.g. data lost during image compression, effects due to variation in light intensity across different LF view). Hence, every corner point in  $P_0$  were compared with  $P_1$  to find the potential match with minimum Euclidean distance using Equation 6.4 and Equation 6.5.

$$P(x, y) = \underset{min}{arg} \left| \sqrt{(P_{0x} - P_{1x})^2 + (P_{0y} - P_{1y})^2} \right| \quad 6.4$$

$$disp_x = P(x) - P_{0x} \quad 6.5$$

$$disp_y = P(y) - P_{0y}$$

Figure 6.12 demonstrates one of the results generated using multiple checkerboards to calculate the disparity data between perspective views (image file 01 to 06, with respect to 00). The corner points detected on all three checkerboards are highlighted with one corner point shown with an enlarged view, in all three checkerboard object planes. It can be observed in the enlarged views that; the disparity generated is in a hexagonal shape with orientation resembling the results shown in Figure 6.8 (bottom right). All enlarged views have been manually resized to fit within the image, but it can be observed that the enlarged view of CB1 consists of more pixel density within the hexagonal grid highlighted when compared to the rest of the enlarged views. The pixel density can be represented as the disparity in the  $x$  and  $y$ -axes. The checkerboard CB3 was selected as the focus plane using the TTF feature with objects CB1, CB2 and CB3 at distances 200 mm, 650 mm and 800 mm respectively. The disparity estimated was recorded against the focus plane distance (TTF) from the camera (CB3 distance in this instance) and the procedure was repeated to obtain five readings with the same focus plane selected manually before capturing LF data.

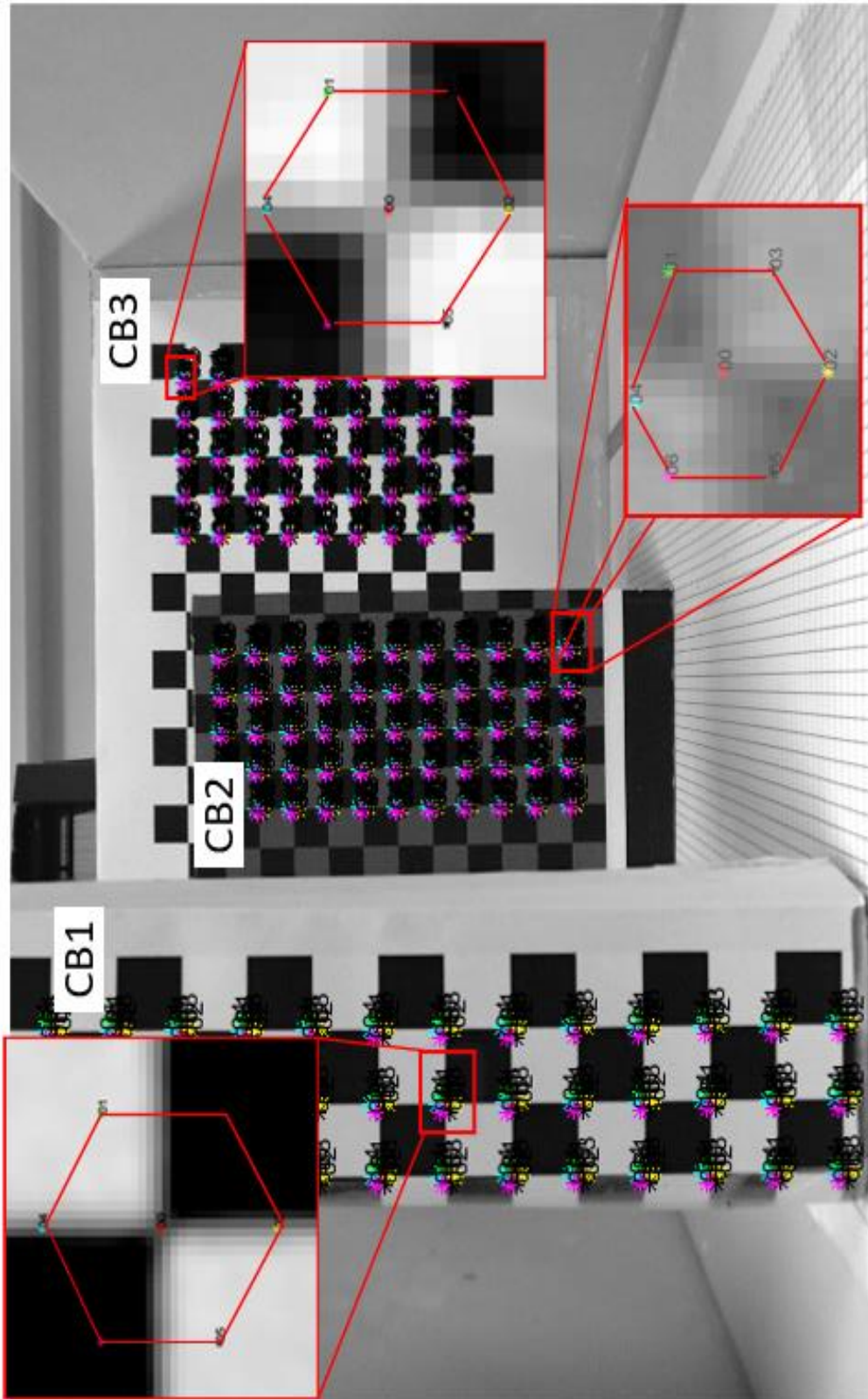


Figure 6.12 Representation of multiple checkerboards used to select the user plane and show one of the experimental results of finding the disparity between multiple checkerboards

The checkerboards were placed in a step fashion (shown in Figure 6.11) to capture three checkerboards within the field of view of the Lytro cameras. This arrangement ensures the visibility of checkerboard corners of all three planes allowing for the straightforward change of focus plane before capturing the LF data between CB1 to CB3 planes. Since the field-of-view of Lytro-I generation is less compared to Illum, only two objects were placed to record the base disparity of the Lytro-I camera.

The procedure of capturing data for the base disparity was to place the checkerboard objects at different pre-recorded distances from the cameras with care was taken to make checkerboard objects parallel to the camera plane. The grid marker feature was activated within the Lytro cameras (Illum), to ensure the checkerboard grid matched up to the grid shown on cameras display (i.e. by activating grid feature, the Lytro camera display show parallel lines/grids). Initially, the plane of focus selected using the TTF feature was at CB1, while CB2 was moved 50 mm steps towards CB3.

This procedure was followed until CB2 reached the minimum distance with CB3, and LF data was captured for every change in position. The change in step distances was carried out back and forth until the image set at every location was equal to five. This gave a measure of repeatability in disparity estimation in the calculation. This procedure was then carried with CB2 and CB3 as the user selected plane and corresponding image sets being recorded.

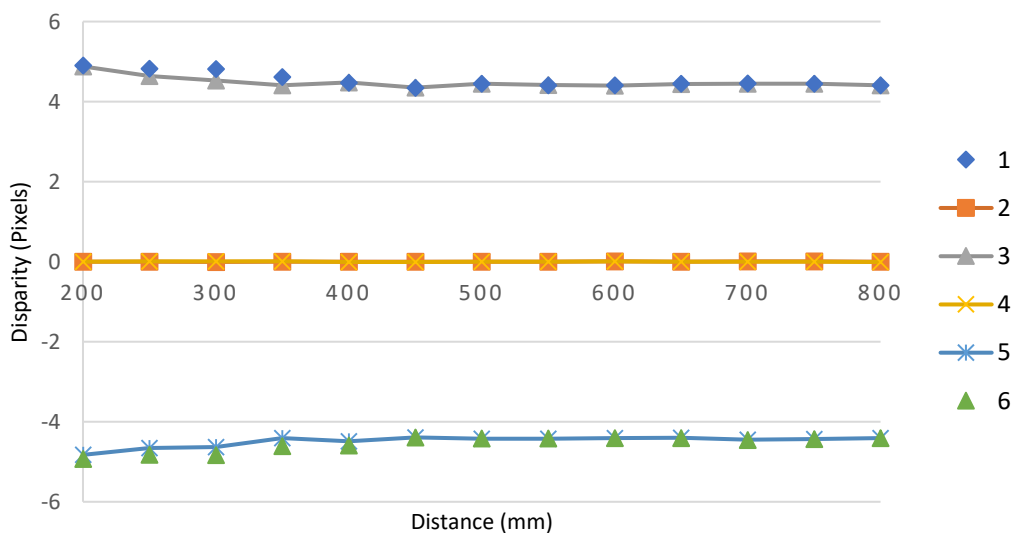


Figure 6.13 Disparity estimation at the focus plane in the  $X$  axis of all perspective views (Illum camera)

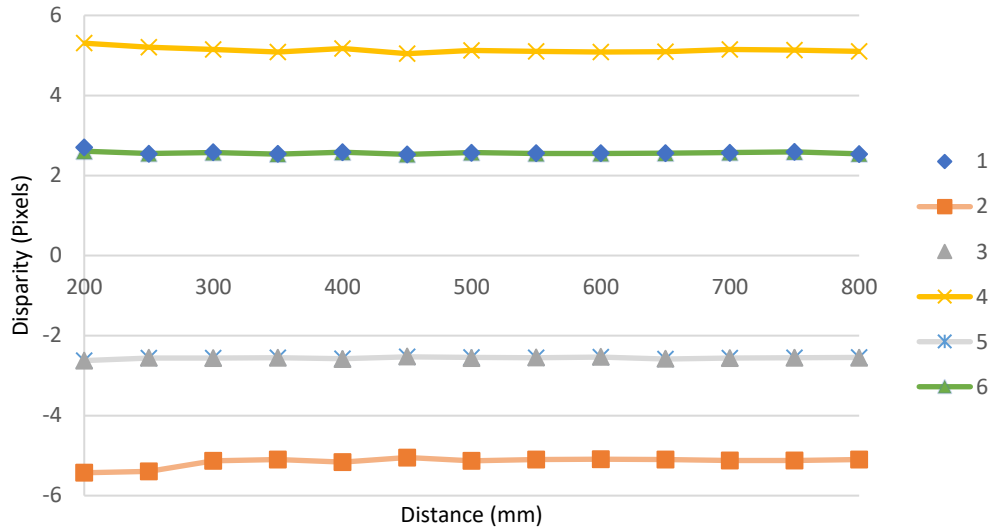


Figure 6.14 Disparity estimation at the focus plane in the *Y* axis of all perspective views (Illum camera)

Figure 6.13 and Figure 6.14 represent the results generated for the base disparity calculation experiment conducted for different TTF planes selected by the user at different distances from the Lytro camera (Illum). In these figures, the disparity calculated for the perspective views (images) with respect to the central view 00 are marked with numbers 1 to 6. The base disparity calculated for all three checkerboards were combined to generate the resulting graphs (Figure 6.13 and Figure 6.14). It can be observed that the base disparity of views 1 and 3 are approximately +4.25 pixels while views 5 and 6 are -4.25 pixels since these views are on the right and left the side of the central view respectively in the *X*-axis (see Figure 6.8 bottom right). While views 2 and 4 are on the bottom and the top of the central view, and hence have base disparity values close to 0 pixels. In contrast, the base disparity estimation in the *Y*-axis shows +5.2 and -5.2 pixels disparity for the views 4 and 2 respectively since they are on the top and bottom of the perspective orientation chart. The views 1 and 6 have a base disparity of +2.8 pixels, and views 3 and 5 have -2.8 pixels disparity. It can be noted that all the views have base disparity values greater than 0 pixels in the *Y* axis since no views fall exactly aligned with a central view.

The base disparity estimation graph indicates the base values in the *X* and *Y* axes that remain consistent with the change in focus plane from 200 mm to 800 mm. It can also be noted that the base values measured close to cameras in the range of 200 mm to 500 mm are inconsistent within  $\pm 0.2$  to  $\pm 0.4$  pixels (see Figure 6.13). But, these values become consistent at greater distances from the cameras (distances greater than 500

mm). A similar response can be seen in the base disparity map for the *Y*-axis, where the views with higher disparity values at initial distances (200 mm to 300 mm) have lower settling time (views 2 and 4). These disparity values are consistent when the focus plane is selected by the user while capturing LF data indicating a reference value in the disparity map generated using the stereo-vision method. If the absolute distance of the user-selected plane is known, then a complete disparity map can be converted into absolute SI unit scale. Thereby reducing the time to calibrate the entire work volume with the depth map generated by the LDS [95]. The calculated base disparity values reduce the amount of time invested to generate a reference calibration map and limited work volume issues [113]. Table 6-2 represents the base disparity values calculated for all six perspectives views of the Lytro Illum camera.

The nature of disparity values obtained from the Lytro-I camera for the same experiments conducted with the Illum cameras resulted with highly variable values (there was no data consistency, with poor repeatability in the disparity values and with a large magnitude of variation to fit on the equivalent graph). Hence no graph/table is defined for Lytro-I cameras. Later in the chapter, the reason behind the lack of fixed base disparity values are explained.

Table 6-2 Base disparity of all perspective views for the Lytro camera - Illum

Perspective views	Disparity in <i>X</i> -axis (pixels)	Disparity in <i>Y</i> -axis (pixels)
01	4.25	2.5
02	0.0	-5.2
03	4.25	-2.5
04	0.0	5.2
05	-4.25	-2.5
06	-4.25	2.5

## 6.6. Quality of complex lens systems

Since the TTF feature is very important for the selection of the focus plane during calculating the base disparity and in applications using this method, it becomes important to determine how good the Lytro cameras can differentiate two planes with high contrast involved. Similar to conventional cameras, the Lytro cameras use contrast detection algorithms to select a plane to be focused by the lens system by adjusting the



zoom and focus stepper motors. This was one of the reasons to choose the checkerboard as target objects since they produce high contrast regions, helping the Lytro cameras to select a plane to focus. Furthermore, for the Lytro cameras to focus on the plane selected by the user, the plane must consist of sufficient contrast to differentiate any other surface within the camera's depth of field for a given lens setup. Hence, using the TTF feature in low contrast regions results in focusing on the different plane by the Lytro camera rather than the plane selected by the user.

It is important that the plane chosen by the user is parallel to the camera plane since a tilted checkerboard plane produces a number of high contrast regions within a short distance. Thereby user-selected plane and plane selected by the Lytro lens system differ. To ease this process, the distance between all corner points in the checkerboard selected by the user was measured and the mean difference was calculated indicating how good the checkerboard plane was aligned with the camera defined as base plane accuracy ( $b_p$ ). Base plane accuracy is given by Equation 6.6 with  $P$  as the corner points detected in the user selected checkerboard plane. The highest base plane accuracy achievable is 1.0.

$$b_p = 1 - (\text{mean}|P_x - P_{x-i}| + \text{mean}|P_y - P_{y-i}|) \quad 6.6$$

In Figure 6.15, CB1 was selected as the focus plane and the disparity estimation procedure was implemented by changing the location of CB2 from the minimum to maximum distance from CB1 towards CB3. Each line in the graph represents the disparity calculated for the 01-view, with CB1 and CB3 at constant locations whilst CB2 changing positions in steps of 50 mm using the Illum camera. It can be observed that disparity values for CB1 (4.3 to 3.8 pixels) are approximately equal to the base disparity (+4.25 pixels). The disparity value gradually shifted to the opposite axis for any plane after the current focus plane was selected. It can be noticed that the disparity values for CB3 are between -3 to -4 pixels since CB3 is far away from the focus plane. The base accuracy was close to 0.79 and recorded at the single location of CB1, hence not shown in Figure 6.15.

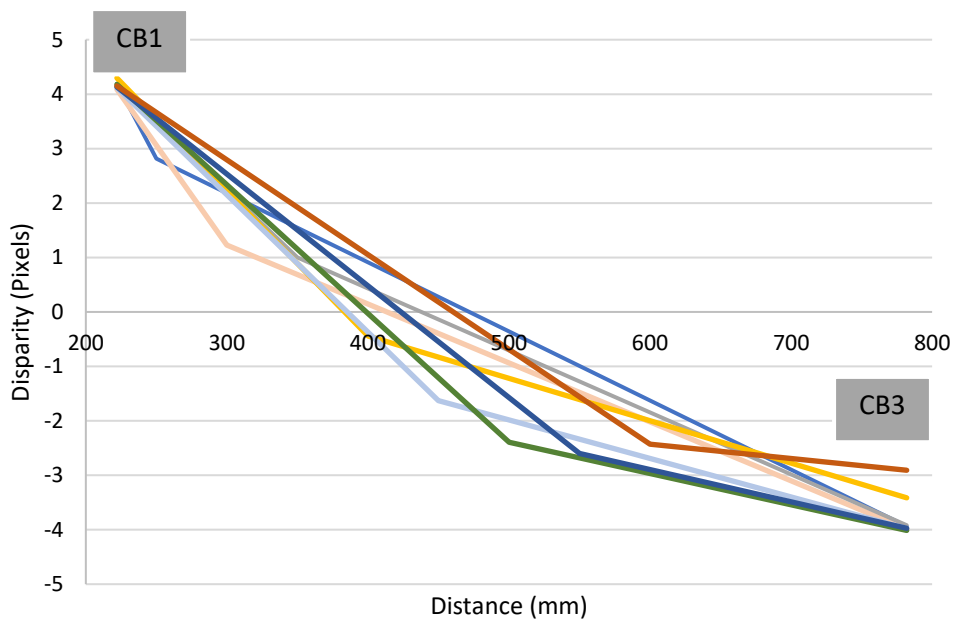


Figure 6.15 Representation of focus plane selection by the Illum camera under user selected CB1 plane using TTF feature and resulting disparity estimation in a X direction (axis) of 01 view

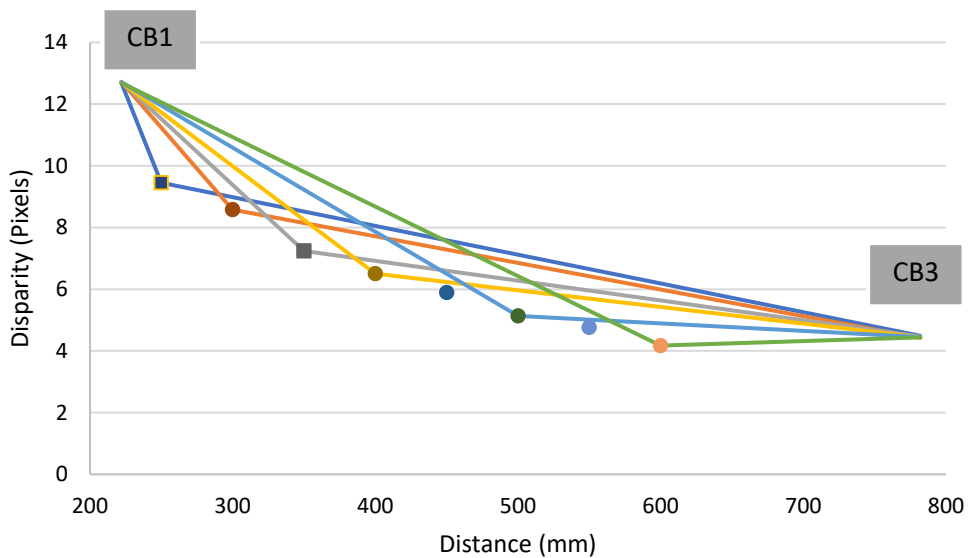


Figure 6.16 Representation of focus plane selection by the Illum camera under user selecting CB3 plane using TTF feature and resulting disparity estimation in a X direction (axis) of 01 view

A similar result can be seen in Figure 6.16, where the user-selected plane was CB3. The CB3 and CB1 planes remain unchanged during the disparity calculation experiment while CB2 was moved towards CB3 in steps of 50 mm using the Illum camera. Since the focus plane was far away from the camera plane, the disparity for 01-view remains constant at +4.25 pixels, regardless of the displacement of plane CB2. The base plane

accuracy was 0.85 at CB3. It can also be noted in Figure 6.15 and Figure 6.16 that the user selected the plane using the TTF was selected by the cameras CLS resulting in disparity values close to the base disparity values.

In Figure 6.17, a similar procedure used in Figure 6.16 was carried out to calculate the disparity values with the CB2 plane selected as the focus plane. With the CB2 target moving away from the camera plane, the base plane accuracy was calculated at various locations and can be seen in the Figure 6.17. The disparity calculated for CB2 remained close to +4.25 pixels, indicating the user selected plane was focused by the camera. Also, the disparity calculated for CB3 plane gradually shifts towards the opposite axis indicating the zero-shift as seen in Figure 6.15.

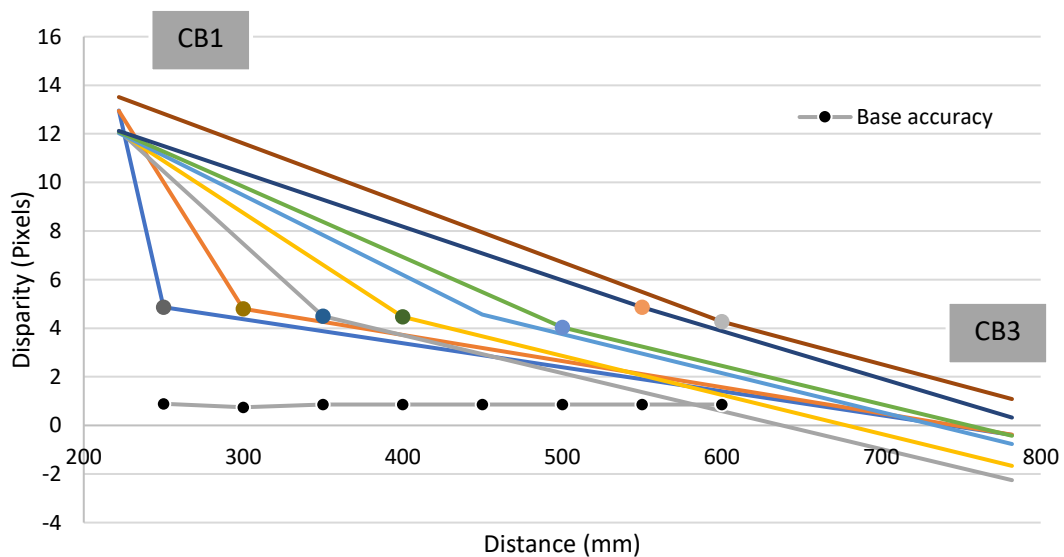


Figure 6.17 Representation of base plane accuracy, focus plane selection by the Illum camera under user selected CB2 plane using TTF feature, and resulting disparity estimation in a X direction (axis) of 01 view

The base disparity experiments with the Illum camera indicate the presence of base disparity at the plane selected by the TTF. Since the base disparity remains close to +4.25 pixels for the 01-view and similarly for the rest of the perspective view as indicated in Table 6-2 regardless of the distance from the camera, the base disparity is used as a reference to generate an absolute depth map. Furthermore, the experiment with the Illum confirms the focus plane matches between CLS and TTF as selected by the user.

These results prove that the base disparity remains constant when the scene is captured using the TTF feature and the disparity changes on either side of the focus plane. To

ease the process of depth estimation, the metric distance between the cameras and the focus plane must be known. Thereby, the entire disparity map can be transformed into an absolute depth map, in this case, an SI unit depth map. The raw image files (.lfr) collected in Chapter 5 were used to generate the relationship between zoom motor steps (*ZS*), focus motor steps (*FS*) and focal length (*FL*) of the Lytro Illum cameras. The data was gathered (see Chapter 5) for a range of 1,000 mm with five sets of images taken at each location varying by 5 mm. The motor values and focal length are extracted metadata obtained by .lfr files using the ExifTool application [114] (ExifTool is available online and is used to decode the formatted text/contents).

In Figure 6.18, the focus motor values drop (steps) initially when the focus plane was close to the camera and gradually increased from 850 to 1,150. The *ZS* and *FS* values are not represented by any units such as millimetres, because these values are obtained from the metadata of the Lytro file and converting these values to other measurands does not add any advantage to the work, since CLS design details are not known and eventually introduces many unknown factors/variables. Also, the motor values are always represented in steps by the Lytro software. It can be noticed that the focus step values gradually remain constant to the change of distance after 500 mm to 1,000 mm, but there was a gradual increase along with increasing distances.

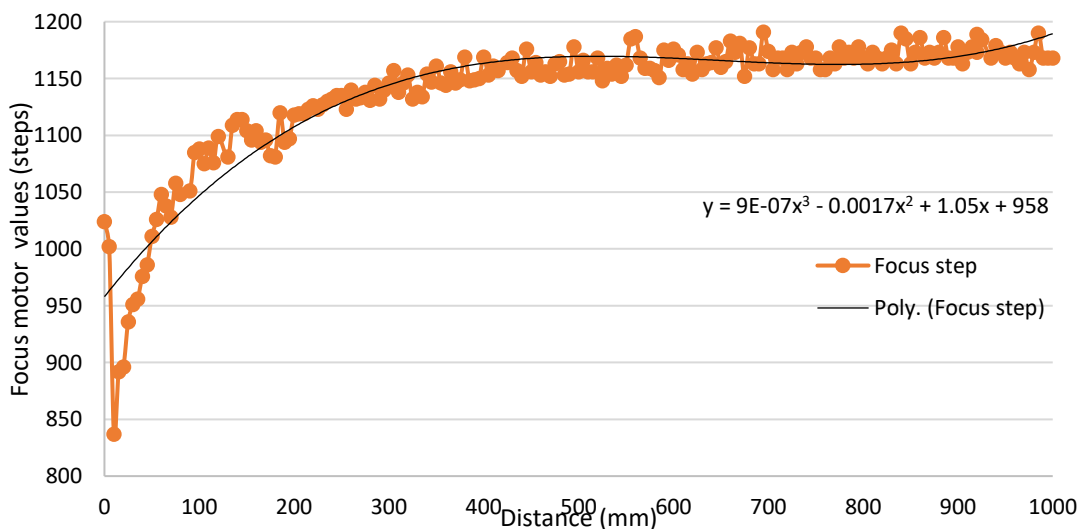


Figure 6.18 Focus motor readings of the Lytro Illum camera focused at different distances using TTF with constant zoom

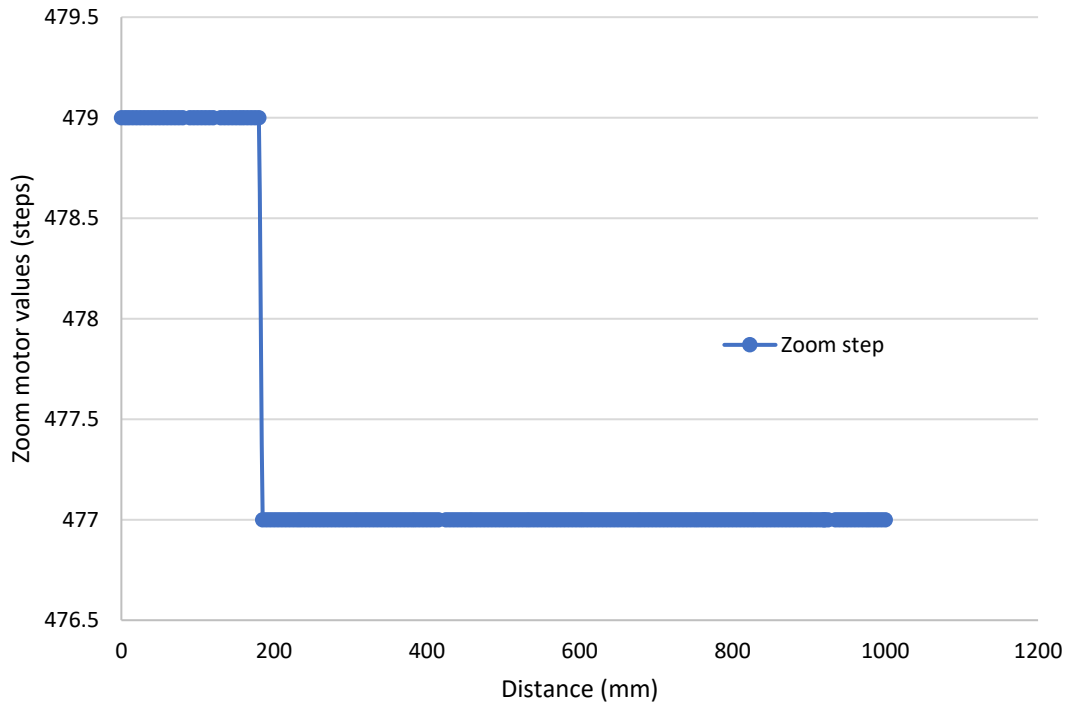


Figure 6.19 Zoom motor readings of the Lytro Illum cameras for objected focused at different distances using TTF feature

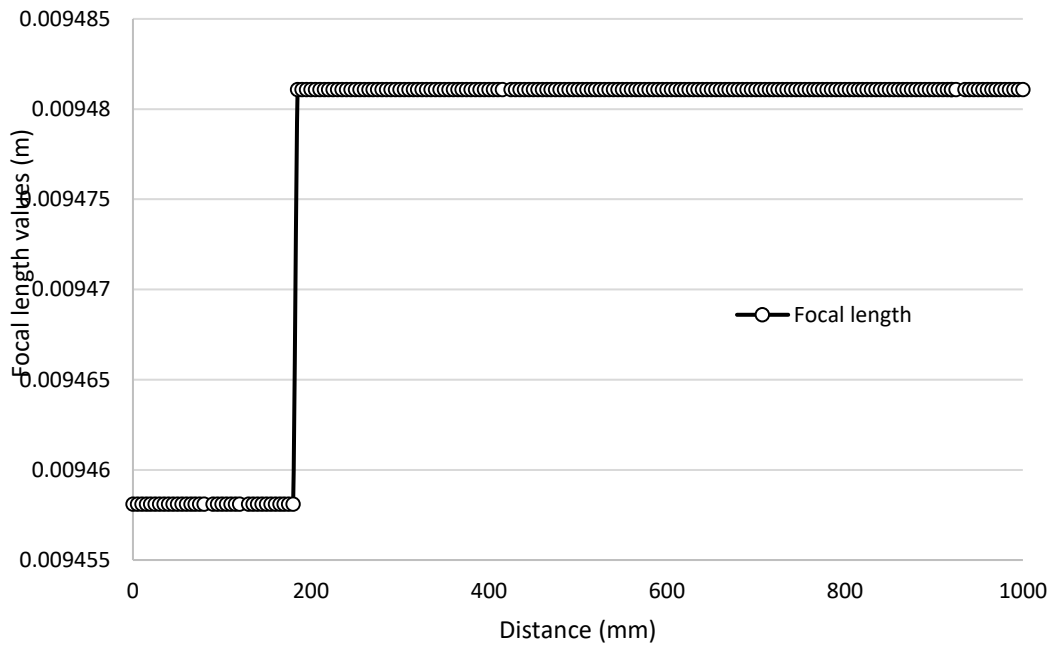


Figure 6.20 Focal length obtained from the Lytro Illum raw files for different values of ZS and FS

The smallest increase in focus step values was 1, with 79 being the highest. Furthermore, it can be noticed in Figure 6.19 and Figure 6.20 that the  $ZS$  and  $FL$  values remain constant with respect to the changing focal plane distance. The  $ZS$  and  $FS$  values are driven by the camera, but the user was responsible for selecting the focus plane using the TTF, and the cameras would select the plane by adjusting motor values. Also, the  $FL$  values are 16-bit values and the changes in  $FL$  values are so small that the plot appears to be constant with changing focus plane values.

To calculate the metric distance of the focus plane, the user needs to extract the  $FS$ ,  $ZS$  and  $FL$  values from the Lytro raw files (.lfr). The metric focus plane distance was found by comparing the values from the raw metadata with the corresponding graph, given by Equation 6.7, where  $\Omega$  is the intersection of data.

$$Z = (FL \cap ZS) \cap (FS) \tag{6.7}$$

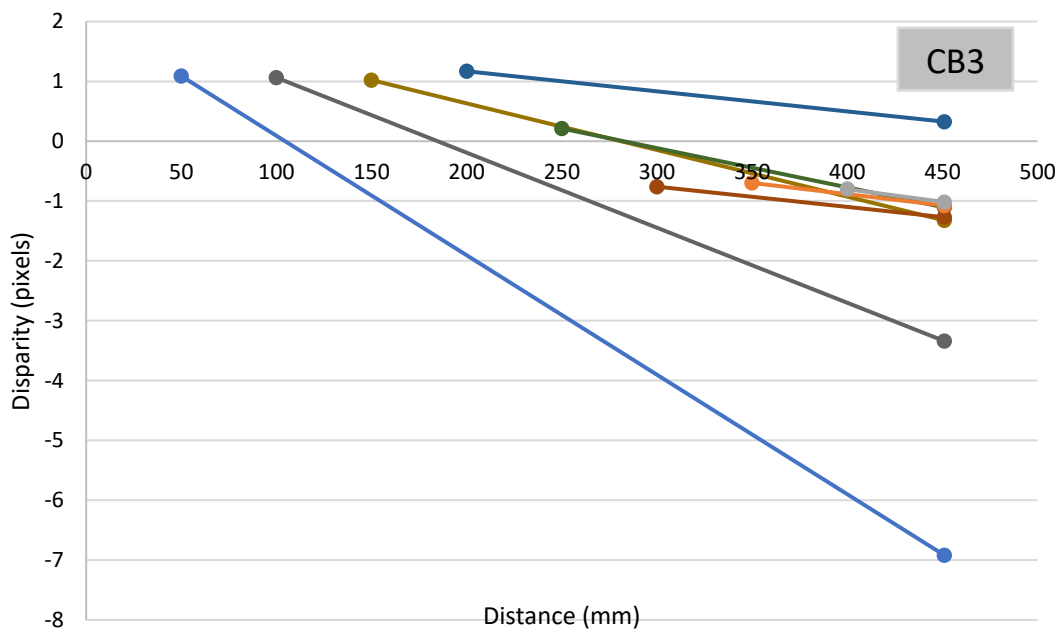


Figure 6.21 Representation of disparity values generated for the user-selected plane (CB1 plane) at different distances from the Lytro -I generation camera (disparity estimation in a  $X$  direction (axis) for 01 view)

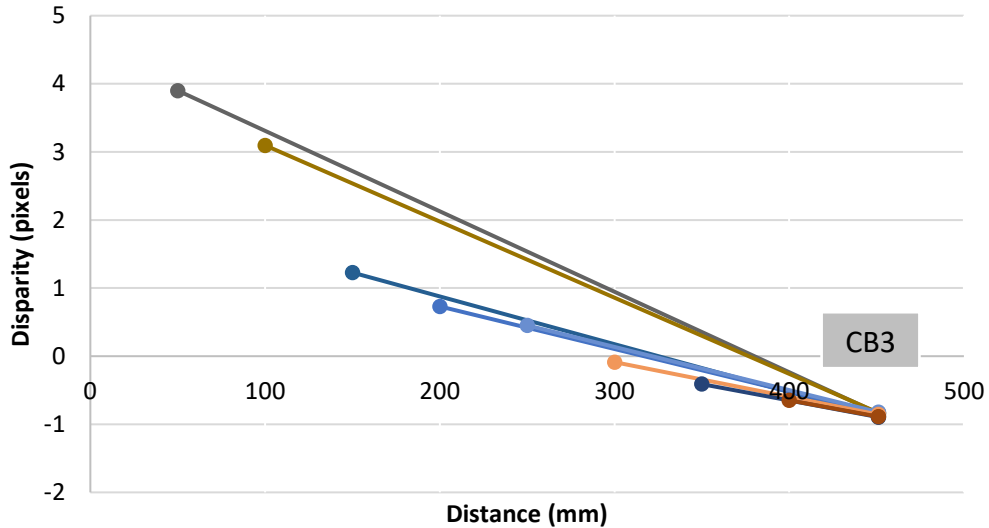


Figure 6.22 Representation of disparity values generated for the user-selected plane (CB3 plane) at different distances from the Lytro -I generation camera (disparity estimation in a  $X$  direction (axis) for 01 view)

The Lytro-I generation cameras have a different setup of CLS compared to the Illum. A 2-piece lens system (zoom lens + focus lens) controlled by a stepper motor forms the lens assembly of the Lytro-I generation cameras. The results are shown in Figure 6.21 and Figure 6.22 define the poor performance of the Lytro-I generation CLS in comparison with the Illum. Figure 6.21 represents the outcome of base disparity experiments showing disparity values at a different distance of CB1 from Lytro-I generation camera. Disparity values remain approximately at +1.0 pixels from 50 mm to 200 mm and then drops close to 0.0 for any distance greater than 300 mm. Furthermore, in Figure 6.22 the TTF selected plane was CB3 and CB1 was displaced in steps of 50 mm towards CB3. The disparity values are close to -1.0 pixels for CB3 plane at 450 mm but a shift of 180 degrees occurs before the focus plane. This effect was also seen in the rest of the views, 02 to 06, which was due to the mismatch between a user-selected focus plane and camera selected plane. It was also noticed that a disparity of +1.0 was stable until 300 mm, which is the range identified in Chapter 5 as an Active Zone for Lytro-I generation cameras. The distances greater than 300 mm were identified as Inactive Zone and a -1.0 disparity support this claim. For this reason, fixed base disparity values do not exist for the Lytro-I generation cameras, however, the rate of change in disparity values can be used for calculating absolute depth.

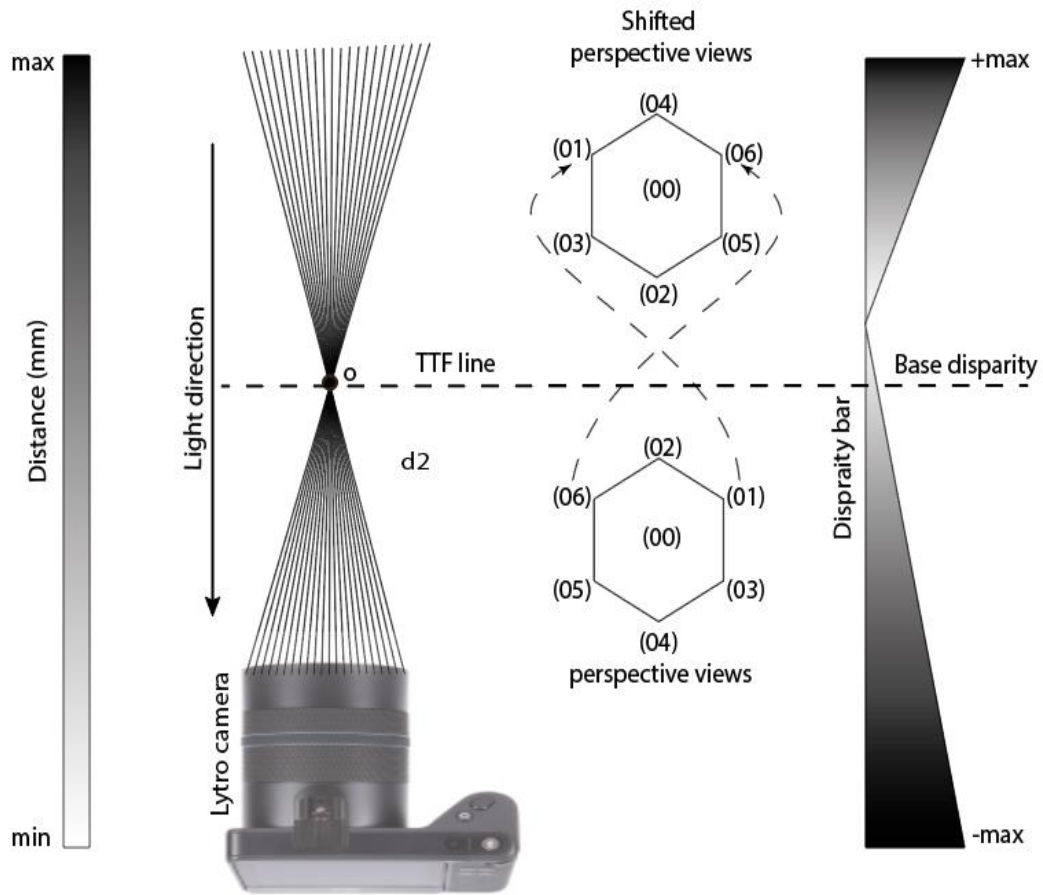


Figure 6.23 Metric depth calculation method overview represented along with disparity bar, TTF line, perspective views and shifted views

The overall method of obtaining the perspective disparity in SI metric scale is shown in Figure 6.23, where the Lytro camera is shown with the light rays emerging from a point source  $o$  selected using the TTF feature. The light rays not only collect the information from the point source but also gather the information that falls in the path between the cameras and the point source  $o$ . The disparity at the point source selected using the TTF will be equal to the base disparity of the respective views (see Table 6-2). Hence the disparity bar on either side of the TTF line is not of the same length. Also, the light rays diverge (pinhole effect) behind the TTF line, thereby changing the sign of the disparity values. This process is represented by highlighting the flipping action of views 06 and 01. Since the base disparity is calculated between the camera photosensor and the TTF selected object, the resulting disparity is measured from the photosensor location rather than the camera exit pupil or the main lens.



## 6.7. Depth estimation in absolute scale

The results generated in Section 6.6 are used to estimate the disparity of a given scene in metric units. For a dense stereo matching, two criteria need to be fulfilled. Firstly, the scene captured by both the cameras should contain some common area or scene to achieve disparity and secondly, the epipolar lines for both images should match thereby easing the correspondence matching. The advantage of using perspective images generated by the Lytro cameras that employ the MLA for capturing LF data is that the images are calibrated by the LDS for lens distortion and all seven images are in the same plane thereby rectification of the stereo-pair image is not necessary. Since the perspective images are arranged in a hexagonal shape around the central view (Figure 6.8 (bottom right)), row shifting of the perspective views is necessary before the images are used to generate disparity. The row-shifting will be applied only for perspective images 01 to 06, to match the pixels of central view. Figure 6.24 shows the results of disparity estimation with and without the row-shifting process, the disparity estimation was carried out with a perspective view 01 as the right image and central view 00 as the left image of the stereo-pair system (see Appendix 10.3). The holes/ missing data in Figure 6.24b is due to the lack of corresponding/matching pairs in the images used for calculating disparity. This result can be enhanced by adopting a row-shifting method which leads to in better alignment of images for finding matching pixel pairs, as shown in Figure 6.24c.

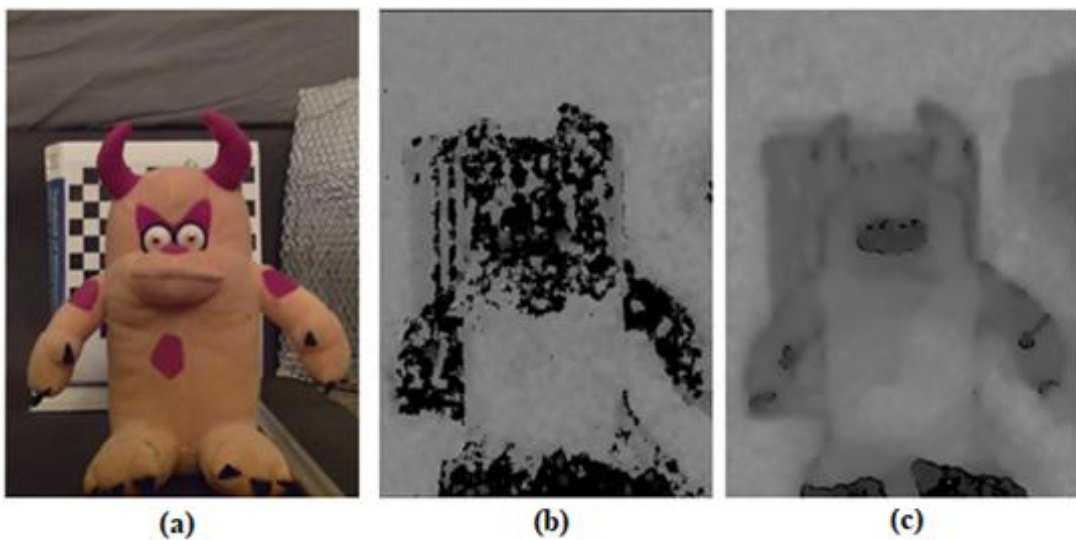


Figure 6.24 (a) Disparity estimation of a scene captured using Lytro Illum, (b) results for images without row-shifting and (c) the result obtained for images with row shifting

$$I = \begin{cases} I_o(1:y - b_{dy}, 1:x) & \text{if } (b_{dy} > 0) \\ I_o(b_{dy}:y, 1:x) & \text{if } (b_{dy} < 0) \end{cases} \quad 6.8$$

Row-shifting was applied to all perspective images using the base disparity data obtained for individual Lytro cameras. The disparity data indicates the off-set involved in each perspective image with respect to the central image in the  $X$  and  $Y$  axes. The Lytro family of cameras exhibit two different behaviours with respect to the focus plane selection. The Illum camera selects the plane as desired by the user as the focus plane by adjusting the CLS efficiently as seen in Figure 6.15 to Figure 6.17, thereby generating a fixed base disparity ( $b_d$ ) at the focus plane. In contrast, the Lytro-I generation fail to adjust the lens system to focus on the user selected plane when the subject (target) is far from the cameras, or in the presence of high contrast subjects within the camera's focus range, as shown in Figure 6.21 and Figure 6.22. Hence to make use of these pros and cons of the Lytro family of cameras, two methods are proposed in this work.

---

**Algorithm: Absolute depth from stereo-view data of Lytro cameras**

---

**1:** Load base disparity ( $b_{dx}$ ), focal plane distance ( $Z_{mm}$ )

Method-1: use values from Table 6-2 and use Equation 6.9 to calculate  $Z_{mm}$

Method-2: use Equation 6.4 and 6.5 on base images and the user given  $Z_{mm}$

**2:** Calculate the disparity map ( $d_{map}$ ) between 01 to 06 views with a central view (00)

**3:** Vectorise  $d_{map}$  into  $d_{val(1..i)}$  with  $i$  elements

**4:** Calculate the absolute distance ( $d_{mm}$ ) for  $n$  elements

**for** ( $i = 1:n$ )

Case 1: ( $d_{val} \leq b_{dx}$ )

$$val = (|d_{val}| * Z_{mm} / |b_{dx}|)$$

$$d_{mm} = (b_{dx} - |val - b_{dx}|)$$

Case 2: ( $d_{val} > b_{dx}$ ) & ( $d_{val} < 0$ )

$$val = (|d_{val}| * Z_{mm} / |b_{dx}|)$$

$$d_{mm} = (b_{dx} + |val - b_{dx}|)$$

Case 3: ( $d_{val} > 0$ )

$$val_{min} = (1e - 4 * Z_{mm} / |b_{dx}|)$$

$$val = b_{dx} + |val_{min} - b_{dx}|$$

$$d_{mm} = (|b_{dx}| * Z_{mm} / |d_{val}|) + val$$

**end**

**5:** Reshape  $d_{mm}$  to match  $d_{map}$  size to generate an absolute depth map

**6: for** ( $i = 1:p_v$ )

    Calculate average depth map using  $d_{mm}$  of all views

**end**

---

Figure 6.25 Algorithm for calculating absolute depth data using base disparity estimation

In method-1 ( $m1$ ), using the base disparity and corresponding absolute distance of the focus plane generated using  $FS$ ,  $ZS$  and  $FL$  for the Illum camera, to generate an absolute depth map. In this method, targets are selected using the TTF and disparity maps are transformed into absolute depth data (suitable only for the Illum cameras). While in method-2 ( $m2$ ), a checkerboard is placed in front of the camera at a known distance and the TTF is used to select the checkerboard as the focus plane. This setup provides a known reference disparity along with the metric distance to camera plane. Using these values, the disparity map is transformed into absolute depth data, however, the reference values are not constant unlike the base disparity values. Method-2 is suitable for the Lytro-I generation and Illum cameras.

The algorithm generates the absolute depth map for both versions of Lytro cameras depending on the type of input data. A semi-global matching technique was used to calculate the disparity between two stereo-pair images [115] as shown in step 2 of the algorithm (Figure 6.25). One of the disparity estimation results calculated from the Illum camera is represented in Figure 6.26. It can be observed the disparity values range between +6 and -10 pixels, indicating that the disparity value increase towards positive values for an object far from the focus plane while decreasing disparity values (negative) for objects close to the camera. Since Figure 6.26 represent disparity data of the Illum camera -4.25 disparity is noticeable in the focus plane, i.e. the closest checkerboard is the focus plane selected by the user. Since, the disparity is estimated by comparing pixels of the right image with respect to the left image, the sign of the disparity is reversed.

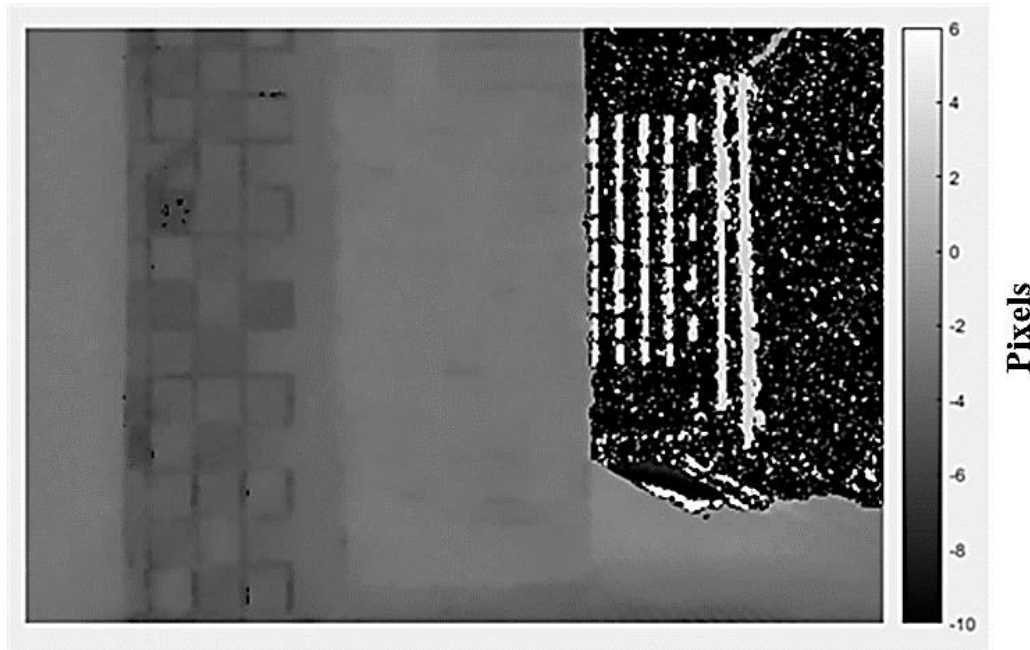


Figure 6.26 Disparity generated between 00 and 01 views

Using the base disparity as indicated from Table 6-2 for method-1, or the user-generated disparity value for method-2, are used to generate absolute distance using steps 3 to 6 of the algorithm (Figure 6.25). The absolute depth results are shown in Figure 6.27 and Figure 6.28, where the top row represents the RGB scene captured using the Lytro cameras and corresponding relative depth generated by the LDS respectively. Further, it can be noticed that the relative depth map and absolute depth map have a close resemblance because the user has no control over the RGB perspective views generated by the LDS. The scene data used by the LDS and this algorithm to calculate disparity are the same, and hence the initial disparity map looks similar to the relative depth map from the LDS. The absolute depth results completely depend on the base disparity and the  $Z_{mm}$  values, hence these values play a key role in this chapter. In this method, the disparity estimation method is less affected by illumination variation, which was one of the key factors affecting the absolute depth map in Chapter 5.

In Figure 6.27, the Illum data was used in method-1 of the algorithm shown in Figure 6.25. Using base disparity values obtained in Section 6.5, all possible sub-aperture images were compared with a central view to generate a disparity map as shown in Figure 6.26. The disparity map was divided into three regions depending on the base disparity value since it is the only known parameter in the entire calculation. The region with disparity values less than -4.25 pixels use step 4 (case 1) equations to represent the

disparity values (in pixel) as absolute depth value (in mm). If the disparity values are greater than the base disparity and less than 0-pixel, step 4 (case 2) and for positive disparity values step 4 (case 3) is used. Using metadata file, values of  $FS$ ,  $ZS$  and  $FL$  were found to be 1103, 477 and 0.0094 m respectively. These values were used to generate  $Z_{mm}$  that was found to be 150 mm (the close match was  $FS = 1104$  for 150 mm and  $FL$  being a 16-bit number the value matched up to 4 decimal point).

Method-1 of this algorithm depends on the absolute  $Z_{mm}$  value and corresponding base disparity value. The depth map generated for the Illum camera shown in Figure 6.27 measured in absolute units (mm), with objects real distance at 150 mm, 200 mm and 451 mm (from camera pupil), while the calculated average distance indicating objects to be at 150.93 mm (neglecting the depth values near grid edges), 196.58 mm and 285.21 mm. In method-2 for the Lytro-I generation cameras, the checkerboard used to generate base disparity was the focus plane at 451 mm with a base accuracy of 0.82. The known distance of objects was at 150 mm, and the resulting absolute results generated using an algorithm (Figure 6.25) were 152.53 mm. The noise in the results are highly noticeable when the object is far away from the Lytro cameras.

It was also noticed that the depth results were close to real-world distances near the focus plane. The greater the object distance from the focus plane the depth results become unreliable. This claim has been demonstrated for the Illum camera in Figure 6.29, but for the Lytro camera-I, many checkerboard planes do not fall within the field-of-view. The Lytro-I gen camera has limited focal length compared to the Illum's higher focal length setting, hence variations in FOV. Therefore, the results are illustrated for the Illum cameras only.



Figure 6.27 Absolute depth generated using method-1 for the Illum cameras

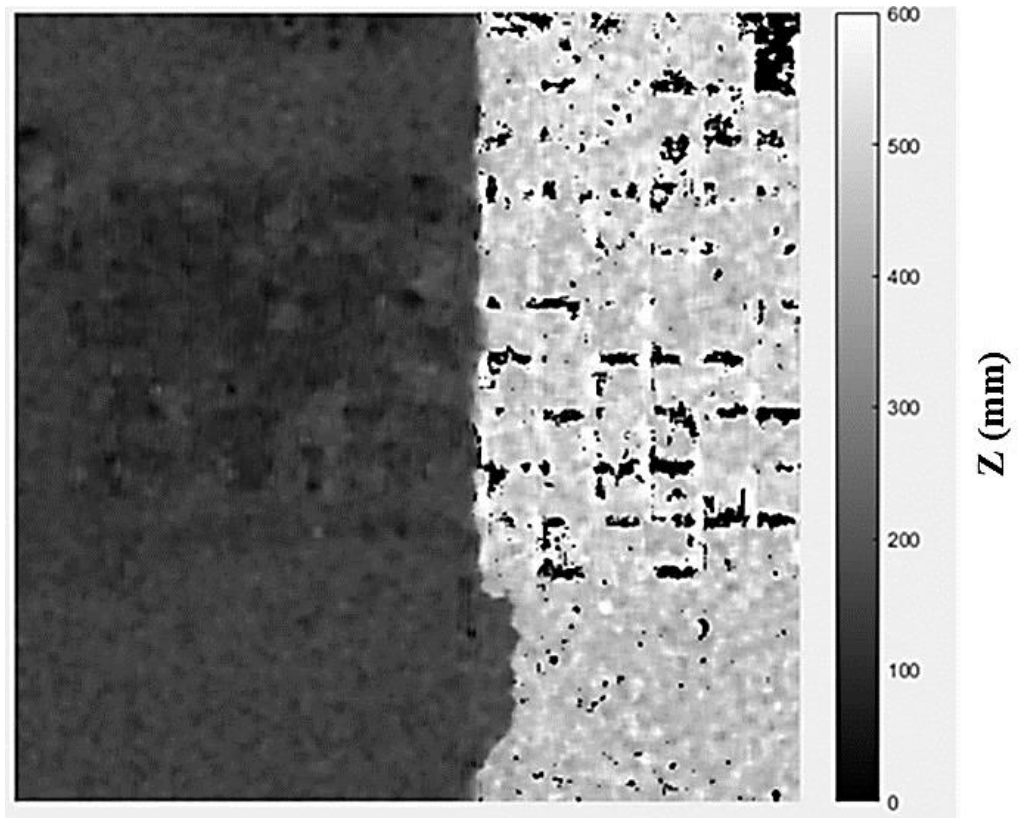
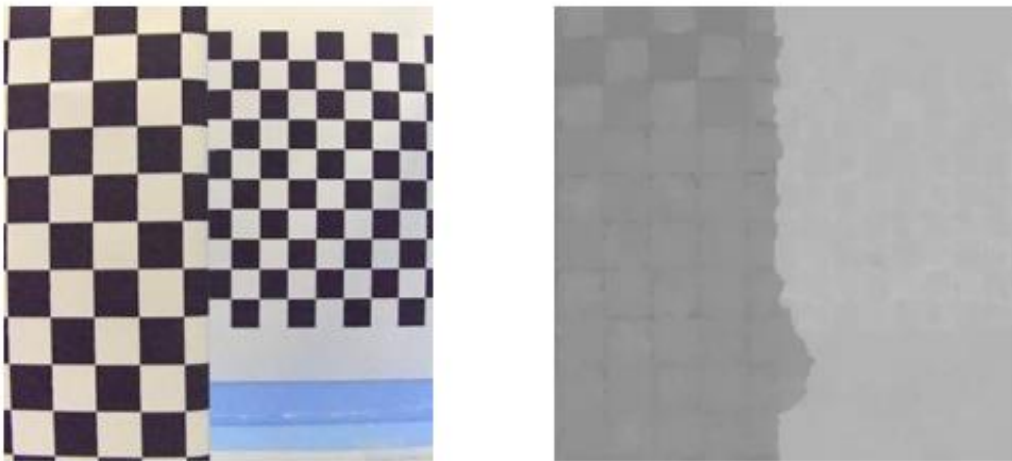


Figure 6.28 Absolute depth generated using method-2 for the Lytro-I generation cameras



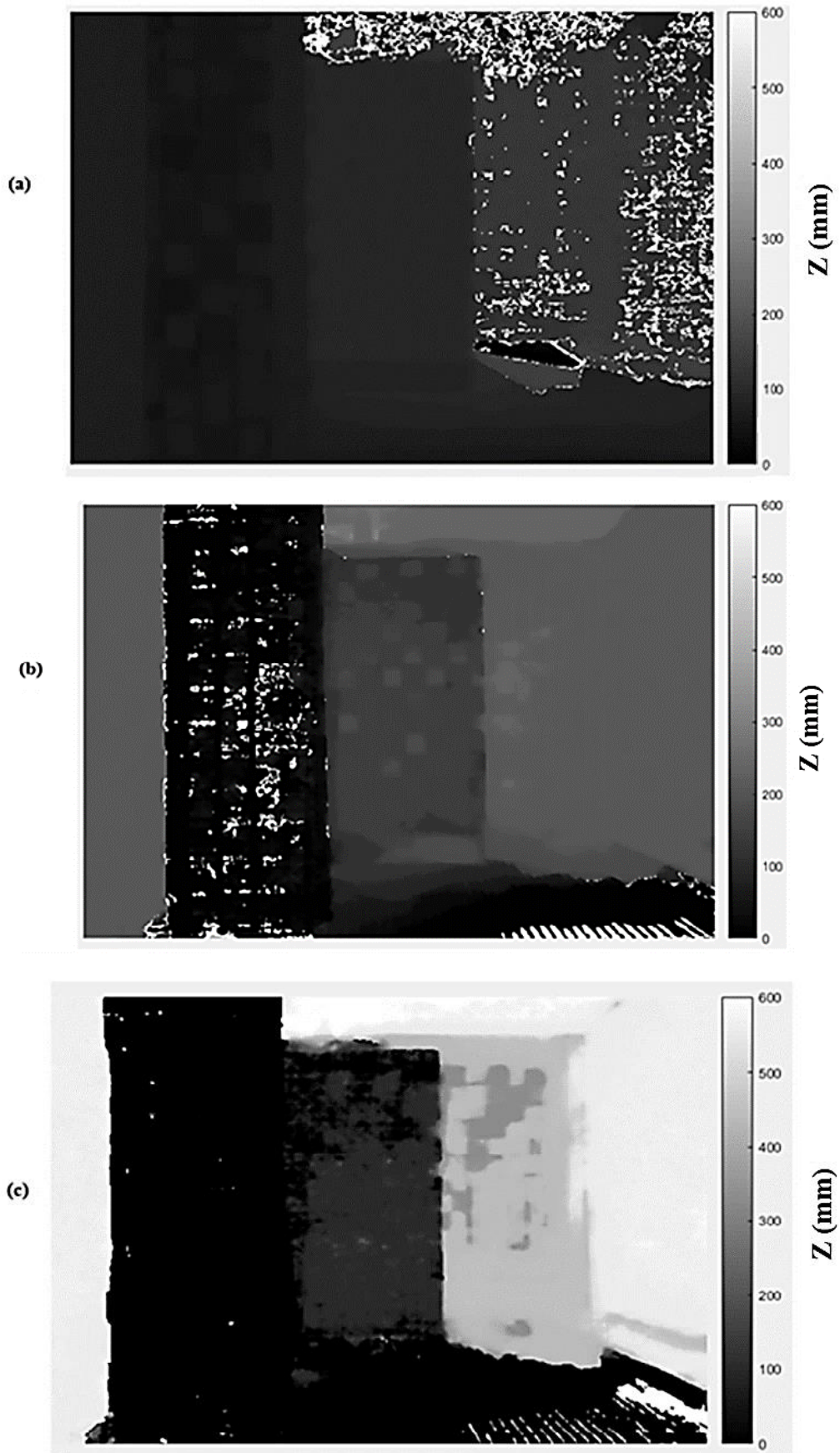


Figure 6.29 work volume illustration using checkerboard plane for TTF at (a) CB1, (b) CB2 and (c) CB3



The images captured for testing quality of the CLS was used to generate depth maps using the algorithm shown in Figure 6.25. The TTF planes are CB1, CB2 and CB3 for three depth maps (a to c) respectively in Figure 6.29. The checkerboards, CB1 to CB3, distances are kept constant at 50 mm, 150 mm and 451 mm while changing focus plane using the TTF. It has to be noted due to the lack of distinctive features in the images (checkerboard images) noise in the disparity images resulted in holes in all depth maps. While using regular scene/objects as target subject (as shown in Figure 6.24), the resulting disparity map and hence the absolute depth maps have consistent depths. Table 6-4, shows the relationship between the focus plane and the accuracy of measurements. The bold values and corresponding checkerboards indicate the focus plane selected by the user. The known depth values and algorithm generated values are compared in Table 6-4 indicating the work volume relationship with the focus plane. It can be noticed that when CB1 acts as the focus plane, depth values of CB1 and CB2 are very close to real distance values (see Table 6-3). The same resemblance can be noticed when CB2 acts as the focus plane. In both these situations, the gap between CB1 and CB2 was approximately 100 mm. With 400 mm gap between CB1 and CB3, and, 300 mm gap between CB2 and CB3, when CB1 and CB2 were focus planes respectively. In contrast, when CB3 was the focus plane, both CB1 and CB2 depth values are far from acceptable values. The distance between different objects in this situation is greater than 300 mm. This suggests the working volume that generates acceptable depth results when targets are less than 100 mm on either side of the focus plane, which indicates a total window of 200 mm for a given focus plane for the Illum camera.

Table 6-4 Accuracy of depth measurement relationship with focus plane

TTF plane	Actual distance (mm)	Calculated distance (mm)	Actual distance (mm)	Calculated distance (mm)	Actual distance (mm)	Calculated distance (mm)
CB1	<b>50</b>	<b>50.9</b>	50	49.2	50	10.9
CB2	150	148.9	<b>150</b>	<b>148.9</b>	150	92.1
CB3	451	200.5	451	280.2	<b>451</b>	<b>454.2</b>

## 6.8. Conclusion

Lytro cameras have an inbuilt feature to select a  $Z$  plane with the help of the TTF. This feature distinguishes the Lytro cameras from other laboratory made the single depth of field LF cameras [49]. Also, the LDS plays an important role in generating necessary images from the raw images captured using the Lytro family of cameras and this has not been completely used in metrological applications. Especially, Chapter 5 demonstrated the use of the Lytro generated depth map in a potential metrological application under predefined working conditions. But, other results generated by the LDS (such as the perspective view) did not contribute to the results of Chapter 5.

The advantages of using the Lytro family of cameras with respect to the single depth of focus LF cameras has been demonstrated in this Chapter. The TTF feature available in both the versions of the Lytro camera was used as a key indicator to define the relative depth map on an absolute scale. The identification of TTF feature in conjunction with generating absolute depth data is the first novelty, as this feature has not been explored by computer vision groups, who typically present relative greyscale results. The simulation results provided by [106] along with the theoretical understanding of the Lytro cameras enable to the identification of the importance of the TTF as explained in Section 6.4, illustrating the major difference between conventional cameras compared to LF cameras to record perspective views.

The novel finding of this work was to identify the relationship between base disparity values and the TTF (see Table 6-2). Figure 6.13 and Figure 6.14 illustrate the variation of the base disparity when targets are very close to the camera in comparison with the faraway targets (within 1,000 mm). Also, the quality and accuracy of the CLS used in the Lytro family of cameras was illustrated and can be considered as proof for variation in performance between Lytro-I generation and Illum cameras, as noticed in Chapter 5 (see section 5.3). The Illum cameras accurately manage to focus on the target set by the user while the Lytro-I generation cameras fail along with zero crossing disparity before the focus plane. The Lytro family of cameras has not been examined over their performance earlier to our group in [95], while evidenced-based performance comparison has been shown in this chapter adding to the novel contribution.

In summary, important points of the Lytro cameras along with few advantages of absolute depth generation methods suggested in this chapter are:

1. Focusing a single plane using multiple cameras (array of cameras) is highly challenging and thereby introduces error. Using Lytro cameras, the selection of the focus plane is straightforward, since the user-selected plane is focused using the CLS.
2. The LDS results are pre-calibrated by the Lytro software and the user need not calibrate the results again if the object distance is greater than 100 mm [95] from the Lytro family of cameras. The calibration model used by LDS successfully overcame any noticeable optical distortion.
3. Using the metadata of Lytro file obtained from LDS, focus plane distance can be calculated in absolute units.
4. The Lytro Illum provide constant base disparity values from which depth calculation is made possible using method-1, while for the Lytro family of cameras method-2 can be used.

The main drawback of the methods suggested in this work is due to the lack of control over the LDS by assigning the algorithm that generates perspective views. Since the work in this chapter is restricted to use the of actual results from the LDS without any manual alteration before depth calculation, the results appear very similar to the relative depth results from LDS. But the results generated by using method-1 and method-2 as suggested in this Chapter provide the depth data in absolute scale (mm) which are one of the novel contributions of this chapter.

A working volume has been suggested in this chapter, that provides acceptable depth data using the depth algorithm. The work volume observes is approximately 200 mm with the focus plane at the centre of the volume. This has been considered as future work that required additional experimentation to generate accurate work volume information. As mentioned in Chapter 5, both Lytro cameras exhibit Active Zone/ work volume and the Illum camera was indicated to have higher work volume of approximately 500 mm. Here the total work volume was identified using the LDS generated results (relative depth map). In contrast, we acquire more control over the depth generation process, compared to Chapter 5 and thereby developing the relationship between TTF and work volume. The advantage of this method is that the

work volume is not static from the camera, but the work volume can be defined to move with the focus plane. Most of the computer vision group working with LF cameras has not defined the term work volume for these cameras, especially for the Lytro cameras (Illum) thereby adding work volume concept to the novelty list of this chapter.

The lack of access to the LDS algorithm prevents from generating all necessary images that result in absolute depth with higher accuracy. This provides a strong reason to have an LF algorithm (reported in Chapter 7) where all features can be manually altered to suit the working condition. This chapter suggested the quality of the Lytro cameras and evidence-based results to carry further research in the areas of LF camera calibration (Chapter 7) and algorithm to generate absolute depth map (Chapter 8).



# 7



## CALIBRATION OF MICROLENS ARRAY BASED CAMERAS

## *Overview*

LF cameras are designed using microlens arrays, that has basic optical parameters such as microlens pitch and focal length. A small number of additional properties can be measured when the MLA is mounted on the photosensor, they are the microimage resolution and lenslets centres (MLA cord). The performance of the LF camera and the quality of the image recorded depend on these parameters. A rigid MLA assembly ensures that these key parameters remain constant regardless of the movement of the camera. These parameters can be used to remove any unwanted distortion introduced due to the lens or mounting issues that affect the captured LF data.

In Chapter 6, the depth measuring technique was limited since the user has less control over lightfield images. To use data captured by low-cost LF cameras and to be independent to process these input data, the user must obtain raw sensor data from these devices and perform lens calibration before any practical application. Chapter 7 mainly describes the steps involved between capturing an image and using the data for applications, such as generating raw sensor data from the Lytro devices, identifying the nature of raw images and performing calibration to remove distortions. Also, the time consumed to generate calibration parameters are also discussed in this chapter.

The information generated from Chapter 7 plays a key role in providing distortion-free images for depth calculation in Chapter 8 and finally gain more control over the low-cost LF cameras

## **7 Calibration of Microlens Array Based Cameras**

### **7.1. Introduction**

Digital cameras are designed to capture a 3D scene on a 2D photosensor. This has been made possible by the lens systems in the camera, that is manufactured for the purpose of bending incoming light. Some cameras may have a group of lenses placed in front of the photosensor to give the users control over focal length, field-of-view and much more [116]. Also, the quality of the glass used to manufacture lenses plays an important role in bending light rays along with other properties such as shape, the radius of curvature and thickness of the lens. Due to design flaws of the lens system or the way these lens units are placed in front of a photosensor, the final image captured by the camera may vary and hence spatial alignment of pixels ( also known as calibration in the field of computer vision [117]) of the lens system is an important step to be followed before using digital images for any engineering applications. Some of the common problems seen in digital cameras are spherical aberrations, chromatic aberrations, and radial distortion due to the lens system [112, 113].

LF cameras are no exception to the distortion as these cameras resemble conventional 2D cameras in design, with a small number of additional optical components [30]. The MLA is one of the additional optical component in the LF cameras that is placed between the main lens and the photosensor. The MLA is a collection of small lens elements which exhibit similar optical properties as the main lens, and thousands of these lenses are embedded in a regular shape to bend light rays towards the optical centre of the individual lens units [88]. A calibration procedure is essential before using the LF camera data in any engineering application to remove any optical distortion and manufacturing flaws. A limited number of calibration methods have been previously published for LF cameras designed using the MLA and these methods follow different routes to achieve 3 axes calibration.

### **7.2. Previous work and motivation**

Previous work on calibrating MLA based LF cameras suggests a 15-parameter camera model that incorporates the methodology from Grossberg and Nayar's work on Raxel Imaging Model and Ray-Based Calibration [66, 114]. This included derivation of a physically based 4D intrinsic matrix and distortion model which related the indices of a



pixel to its corresponding spatial ray. In this method, a practical objective function based on ray reprojection was defined and an optimization framework for carrying out calibration was presented. They also described the method for decoding hexagonal lenslet based plenoptic images without prior knowledge of the camera's parameters and related the resulting images to the camera model.

In [88], the misalignment in the microlens array was taken into account by providing an estimate of the position and orientation of the microlens array and then calculating the calibration parameters using camera settings. A geometrical model was developed to relate the location of a microlens to the location of the focal point on the image sensor to determine the position of the array that best correlated to the known focal centres. In this work, synthetic and experimental calibration images were used to determine the robustness and accuracy of the algorithm. On the account of determining the position and orientation of the microlens array, the geometrical method provided an estimate of these values. By knowing these values a calibration model was generated and later applied to the LF cameras.

Similarly, in [85], the geometric calibration of MLA LF cameras was defined by using line features extracted from raw images directly, instead of sub-aperture images. A projection model based on the thin-lens model and the pinhole model has been applied to line features by which an initial solution of both intrinsic and extrinsic parameters was estimated using linear computation, and refined via a nonlinear optimization. This work also defined the root mean square (RMS) ray re-projection error to be very close to the work of [70], which was 0.0717 and 0.365 for 3.6 mm and 35.1 mm checkerboard square grids (the RMS reprojection error values are the difference between of actual and obtained values, a pixel location in this context).

Most of these existing works describe new methods and approaches for better calibration of MLA based LF cameras but no work has been found that considers using known details of the LF camera in calibration such as; MLA pitch, perspective views, and direction of light rays recorded on the photosensor. Most of the previously published papers describe calibration using a raw 4D LF image with the emphasis on exploring new algorithms but the time is taken to achieve results are typically neglected. Also, some of the existing calibration algorithms generate 2D images of LF data which is insufficient for applications making use of perspective views and post-

focusing features. While generating depth results in absolute scale, calibrated 4D data is necessary for generating image stacks. As a consequence, the research developments illustrated in this chapter describe an approach that is influenced by 2D camera array calibration methods using the pinhole camera concept.

Theoretically, in a multi-camera system, the camera arrangements are made as a way to capture the same scene data from different viewpoints. In this scenario, the distance between a two camera system is in terms of a millimetre (mm) and hence a single calibration method similar to Zhang [121] can be applied to the individual camera and a method similar to [122] can be used for multicamera array calibration.

In contrast, the key parameters generated by the MLA-photosensor (MLA-Ps) assembly such as MLA pitch, microimage resolution and LF resolution play a key role in calibration. With this information, the LF data can be modified into individual camera views where the virtual distance between each camera view is in micrometres ( $\mu\text{m}$ ). So a calibration method similar to the multi-camera system can be used to calibrate LF camera data with known LF camera physical parameters. Hence, for a calibration algorithm to perform independently of input source the raw sensor images must be extracted and later the individual camera views are generated for calibration.

### **7.3. Light field calibration**

The MLA is an important optical element in the LF camera, causing the photosensor to record light intensity along with directional data. The presence of an MLA poses new challenges to calibrate the LF camera since every individual microlens bends incoming light rays towards the respective microlens principal axis, which does not necessarily align (parallel) with the camera optical centre (see Chapter 4, sec 4.2). The straightforward approach is to consider the main lens and each microlens as a lens pair for calibration. But this is computationally expensive, with the generated calibration parameter being of the order of the number of microlens units in the MLA. Also, resolution of the microimage is limited, for example, 10x10 pixels (Lytro I-gen) and 13x13 pixels (Lytro Illum), and noise in the microimages will offset the calibration parameters due to insufficient pixels [123]. The intensity variation in the microlens images increases, when further moved away from the main lens optical axis due to vignetting [86] (the amount of light spread across the diameter of the lens is uneven). Hence, finding multiple feature points in a microimage and hence the correspondence

with microimages of neighbouring microlenses is difficult and time-consuming [118, 119]. In contrast, a conventional camera consists of a complex lens system symmetrical to the optical axis of the camera, hence a pinhole model-based calibration method can be used to calibrate these cameras considering multiple lens assemblies.

In this work, the LF calibration is broadly classified into two sections, calibrating free-floating 2D views and multi stereo lens calibration. In [88] and [85], the MLA cord was detected using a geometrical structure where the optical centre of the microlens was considered as the principal axis of the microlens. In contrast, the geometrical structure for calculating the MLA cord is neglected here and the direction of the light ray is considered, i.e. considering the principal axis as the microlens centre. The MLA cord found using the geometrical structure and light directed methods differ in the order of a few pixels (approximately 3 pixels for the Lytro family of cameras, see Chapter 4 for more information), hence the resulting sub-aperture images may vary. The calibration method described here, overcomes the error in defining the MLA cord and generates calibration parameters, defining the first novel contribution in the LF camera calibration section (see Chapter 4).

The stereo camera calibration procedure considers aligning the epipolar lines of the cameras, thereby the resulting pixel data improvement helps to find the disparity as a computationally inexpensive linear 1D search [111, 120]. To reach good stereo calibration of two (or more) cameras it is important to know the relative positions of cameras with respect to other cameras (in some cases require data with respect to the target object, a checkerboard for example). This can be computed without prior knowledge of camera positions using homography ( $h$ ) techniques and camera parameters (see Appendix 10.6) [126]. If the relative distances of the cameras are known (baseline), the stereo camera calibration parameters can be verified with a known reference parameter. This helps in additionally generating the calibration accuracy value, along with which ray re-projection error values.

LF cameras resemble conventional cameras in lens arrangements, hence they can be considered as multi-stereo camera systems by considering sub-aperture images. Also, they exhibit several advantages over conventional stereo cameras or an array camera calibration with respect to build quality. In a stereo camera system, errors are introduced while aligning the camera plane normal to the optical axis of each camera

and errors are in the order of a few millimetres to tens of millimetres [100]. In contrast, the sub-aperture images provide accurate alignment of the virtual camera plane in the order of micrometres with the lens optical axis. Since the MLA is a thin, rectangular (maybe square) optical element embedded with tiny lens elements that bend light rays towards individual microlens units. The thickness and microlens surface inaccuracy are in the order of micrometres and the constructional accuracy is around 95% to 99% (see Table 7-1) [68, 69, 121] (see Appendix 10.4 and 10.5).

Table 7-1 Tolerance of MLA parameters by different manufacturers

MLA Manufacturer	Height (mm)	Width (mm)	Thickness (mm)	Pitch (mm)	Focal length (mm)	Radius of curvature (mm)
Thorlabs	6.0±0.05	6.0±0.05	1.0±0.05	0.06±0.001	0.24±0.01	0.11±0.01
Edmund Optics	10.0±0.05	10.0±0.05	0.9±0.05	0.3±0.005	4.8±0.05	2.2±0.05

The MLA acts as a rigid lens holder, holding the microlens elements in a fashion such that the optical axis of an MLA unit is parallel to the optical axes of the neighbouring lenses (given MLA constructional accuracy is high). This provides a stable platform for defining the free-floating (2D camera views) images with known camera parameters and optical centres.

The second advantage of the MLA over conventional stereo cameras is the baseline accuracy. With conventional cameras, the baseline plays an important role while calculating the disparity map. Since the baseline is inversely proportional to the disparity map [128], small errors in the baseline value typically result in a large error in the depth results. Reducing the baseline error by adjusting the camera setup is difficult since the camera parameters are unique and differ between cameras along with the added human error. Hence, generating accurate and repeatable stereo setup is difficult with conventional cameras and additionally requires generating new camera calibration parameters with slight modification in the setup. With the MLA based cameras, the baseline problem and generating new camera calibration parameters can be minimised with highly repeatable and accurate baselines in the order of micrometres, along with

single camera calibration for a given camera, under the assumption of known MLA parameters (see chapter 4). This is because the MLA basic parameters and parameters of the MLA with the photosensor do not change with camera movement (e.g. internal changes such as a lens to lens distance or replacing the main lens and external changes such as a change in camera location with respect to a fixed target). The baseline between neighbouring microlenses must be calculated to use absolute depth generating methods illustrated in chapter 6.

Also, as previously identified calculating the disparity error for every microlens is computationally expensive, with the computation time being equal to the number of the microlenses in the MLA. To minimise the effort to find the disparity for every microlens, a virtual lens technique is used that accounts for all directional light rays. The virtual lens method reduces the stereo lens calibration parameters which are equal to the number of the microlenses in the MLA to the pitch of MLA, i.e. to calculate the stereo lens calibration parameters for the Lytro I-gen is approximately 328x328 parameters which are reduced to 10x10 parameters using the virtual lens method.

### **7.3.1. Raw light field image**

In Chapters 5 and 6, the primary source of the LF data used to calculate the absolute depth was generated by the LDS. The LDS uses the raw data captured by the Lytro family of cameras to generate different results such as depth map, perspective views and more. It has been explained in earlier chapters that to use a commercially available LF camera (Lytro) in engineering applications, the software used to generate LF results should be user-friendly, i.e. the user is permitted to access to all software parameters to generate necessary results. Since the LDS does not provide access to understand or alter the software engine, it was considered as Black-Box in Chapters 5 and 6. To gain control over all parameters used to generate the LF results using data captured by the Lytro family of cameras, the raw image file plays an important role. The Lytro company has not provided any features within the LDS to generate the raw image file captured by the Lytro family of cameras.

Since the Lytro family of cameras are considered as the core of this work and the primary source of LF capturing device, generating raw LF data from these cameras was important. It was observed that the raw LF data captured by the Lytro cameras along with the parameters of the camera are stored in .lfr and .lfp format files (metadata) for

the Illum and Lytro-I generation cameras respectively. These Lytro files can be decoded into different components by using the LFToolbox in MATLAB specifically built for extracting raw LF data [129]. LFToolbox version 0.3 was used to generate raw photosensor data from .lfr and .lfp files. The software depends on the Lytro cameras until this stage and the rest of the process are independent of the Lytro camera.

A comparison of the raw LF data generated using LFRReadLFP command (LFToolbox) and one of the sub-aperture images generated using the LDS is shown in Figure 7.1. It can be seen in Figure 7.1 (right) that the image is distortion-free and has single perspective data (see the zoomed portion highlighted by the red and blue box). In contrast, the raw LF image has visible radial distortion with multiple perspective views in the image as shown in Figure 7.1 (left) (multiple views indicate LF data). This indicates the LF image generated using toolbox has not undergone any image processing or calibration steps that the LDS applies while generating images/results. Hence the raw image provided by the toolbox can be considered as an independent LF input source from the Lytro family of cameras. Similarly, any MLA based camera images can be used as an input source for this algorithm. Further, the raw image generated by the LFToolbox must be calibrated to remove radial and any tangential lens distortion (see Chapter 3).

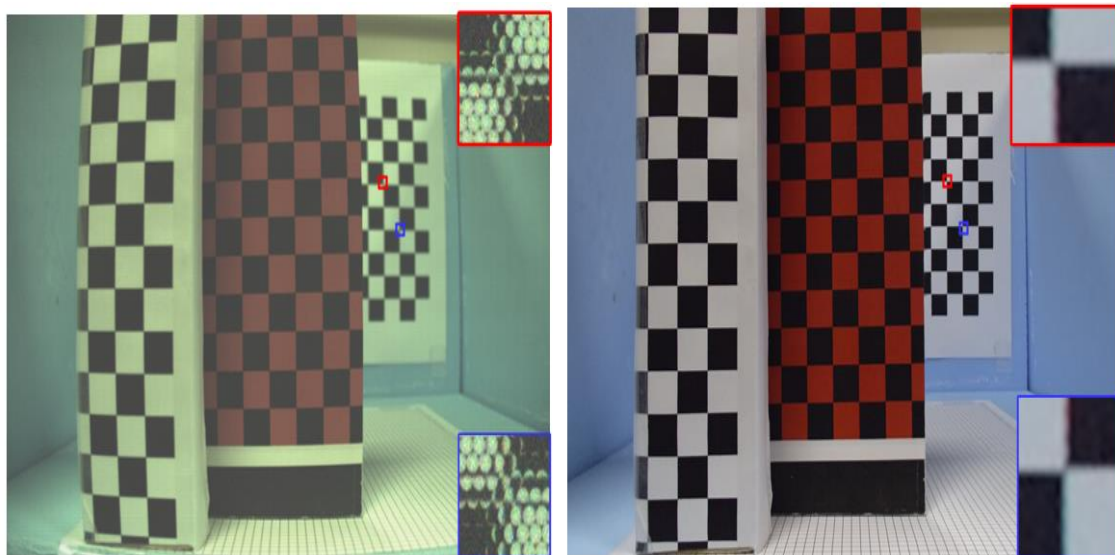


Figure 7.1 LF raw image generated using toolbox (left) versus sub-aperture view generated using LDS (right) of an object at 200 mm from the camera.

Before using the raw LF images of the Lytro family of cameras, the nature of pixel registration must be examined to ease the calibration process. MLA based LF raw data exhibit two optical phenomena while registering the light rays on to the pixels such as; telescopic and binocular views [63, 130]. It becomes important to understand the behaviour of the LF input source (Lytro cameras in this case) for calibration because if the camera exhibits both optical phenomena at the focal plane then the input image must be rearranged with respect to MLA lenslets [66, 78].

### **7.3.2. Dual views of the raw image**

Figure 7.2 and Figure 7.3 illustrates the two views exhibited by the Lytro family of cameras and represent the behaviour at the focal plane, and either side of the focal plane. The same effect was described in Chapter 6 as a zero-crossing for the sub-aperture images generated by the LDS. Figure 7.2 (top row, left) represents a raw image with the focus plane at 200 mm from the camera consisting of three checkerboards arranged in step fashion. Since the checkerboard close to the camera was selected using the TTF, the light rays represent the binocular view between the camera and focal plane, while representing a telescopic view of any region after focus plane. Similarly, Figure 7.2 (top row, right) represents a raw image with the focus plane at 800 mm from the camera and the checkerboard from the camera was selected using the TTF, the light rays represent a binocular view between camera plane and focal plane.

This indicates that the Lytro family of cameras consistently exhibits a binocular view from the camera plane up to the focus plane and the rest with a telescopic view (see Chapter 3). Due to this consistency, image processing for rectifying the raw image is not necessary (since the camera design does the necessary flipping of pixel location). Also, the bottom row of Figure 7.2 represents the enlarged view of the checkerboard corner for a better view of the different views in action. Under the binocular view as we move in any direction, the edge move relative to the MLA centre in the same direction, i.e. in Figure 7.2 (bottom row A) while moving from the black to the white side, the black pixels always stay to the majority side of the microimage (left-hand side of MLA). Similar movements are observed in Figure 7.2 (bottom row a, b and c). while in the telescopic view, the edge always stays on the opposite side of the direction of movement, i.e. the black pixels stay on the minority side (right side of the MLA).

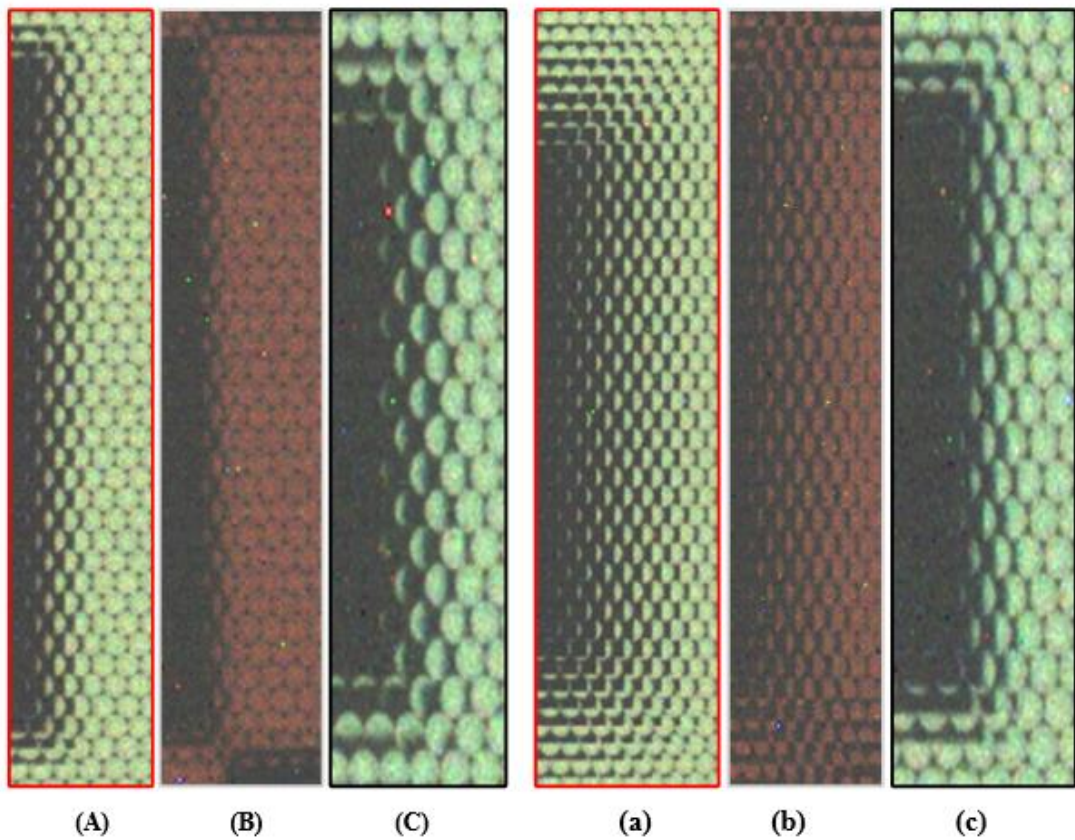
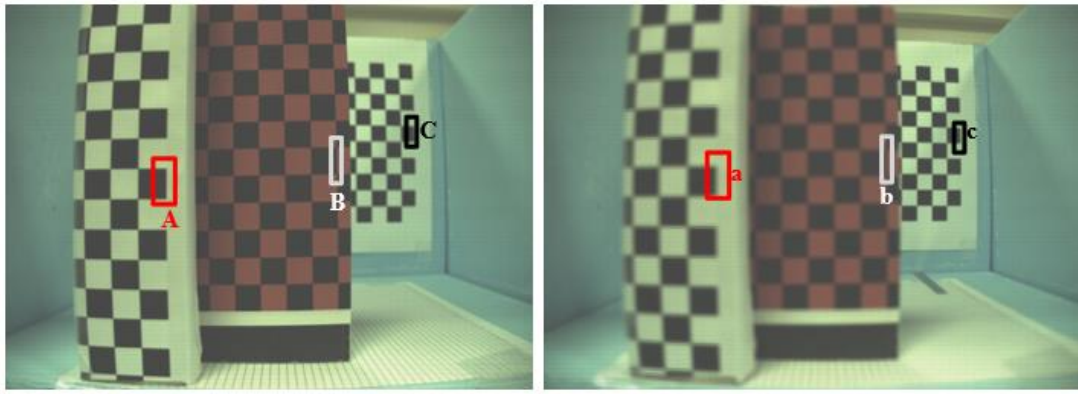


Figure 7.2 Representing telescopic and binocular views by the Lytro family of cameras (Illum) at 200 mm (top row, left) and 800 mm (top row, right) focal plane from the camera

This effect can be observed in Figure 7.2 (bottom row B and C), the out of focus region after the focus plane exhibits telescopic view. In [63], these two optical views were exploited along with some optical adjustments to generate full radiance images with high-resolution image output. The telescopic and binocular views can be compared with real and inverted image effects respectively. Similar views for the Lytro-I are shown in Figure 7.3 (top), while enlarged views show binocular and telescopic views with pixel arrangements.



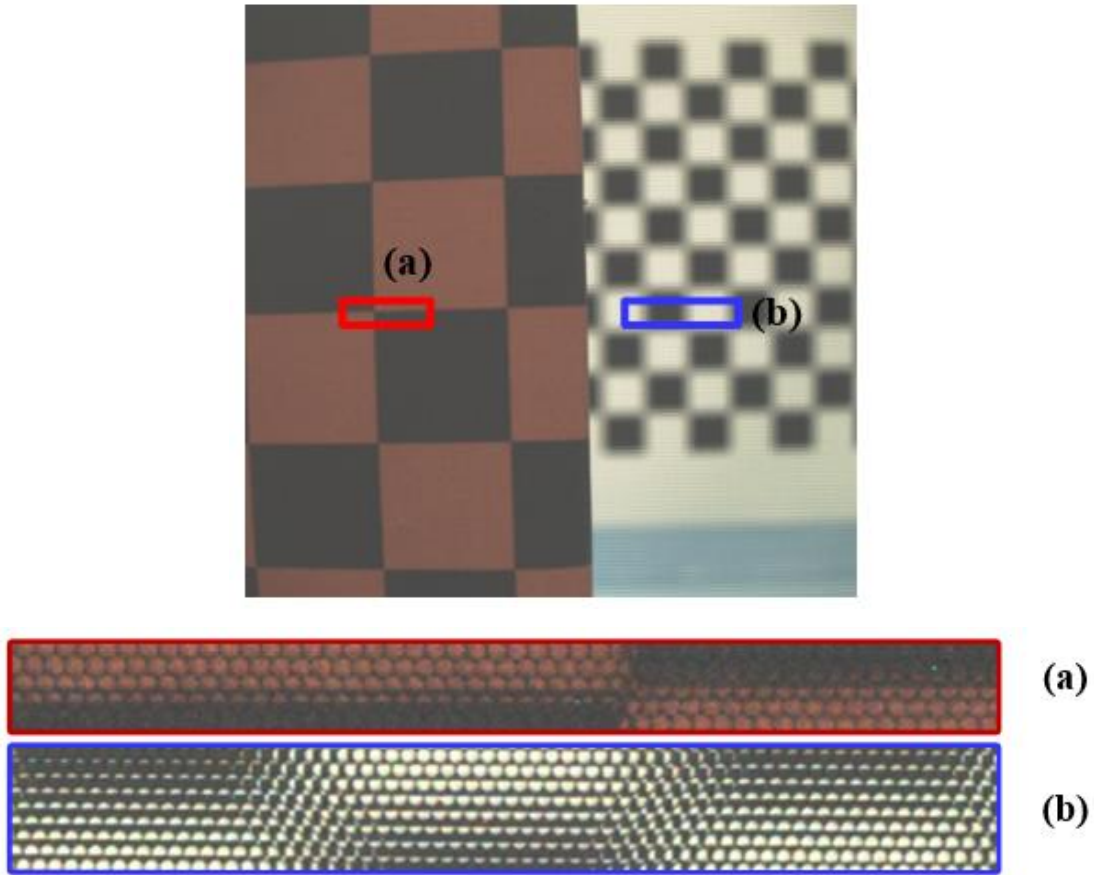


Figure 7.3 Representing telescopic and binocular views by the Lytro I-generation cameras at 200 mm (top) focal plane from the camera and corresponding enlarged views (bottom)

### 7.3.3. Preferences of the calibration image

Calibration of imaging devices is usually performed to remove unwanted distortions that are introduced by the camera optics along with alignment with the photosensor. In the case of conventional cameras, since the final image represents single plane (depths are sandwiched), corner detection is not affected by the position of the target plane (checkerboard) with respect to camera's plane.

In contrast, and as explained earlier, the LF cameras exhibit two different views in the raw images (dual view). Figure 7.4 represents the classification of regions in front of the camera into two different views by considering point object  $o$  as target plane. The TTF defines the separation boundary between the views when activated by the user. When the image used for initial calibration exhibits both views in a single image, corner detection tends to generate invalid results. The corner points detected on the binocular side of the image tend to have positive differences, while on the telescopic image side they show negative differences when compared to a reference point.

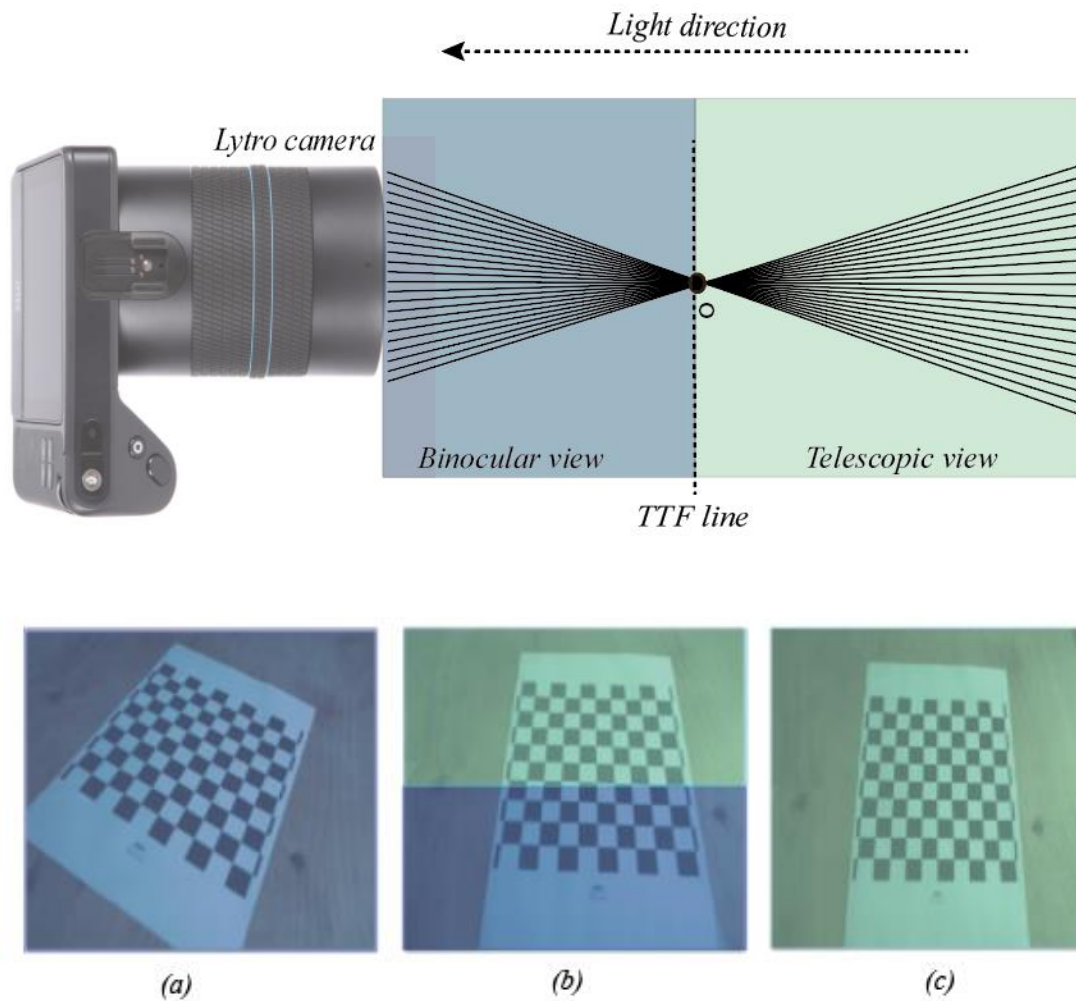


Figure 7.4 Representation of binocular and telescopic view exhibited by Lytro family of cameras defined using a checkerboard image

To avoid a dual view problem in the calibration image set, the input images are manually adjusted to stay in a particular view. This can be accomplished by manually selecting the target plane using the TTF and decreasing the gap between the camera and target plane to ensure the captured image completely stays in the binocular view and vice-versa for the telescopic view by increasing the gap. The resulting images are shown in Figure 7.4 (a) and (c), while Figure 7.4 (b) shows a dual-view effect in the single image.

#### 7.3.4. Transformation of light field data

The raw images of an LF camera are a collection of light rays from different directions of the same scene and arranged according to the shape of the individual lenses in the MLA (see Chapter 3 and 4). For example, the raw LF data of the Lytro family of cameras are arranged in a hexagon pattern since the Lytro cameras (Lytro-I and Illum)

have hexagonal shaped microlens units in the MLA (Chapter 6). The arrangement of light rays on the photosensor depends on the angle at which light rays are reflected from an object surface in the world space represented by  $[X, Y, Z]$ . These light rays pass through the intermediate plane, which is the region between the object and the MLA, consisting of the camera main lens (CLS)  $(s, t)$ . The light rays traveling from the MLA plane towards the photosensor plane are recorded on a particular pixel that depends on the angle at which light rays were reflected from object surface in world space (see Chapter 3, Figure 3.6). The light rays reflected at a higher angle from the object surface with respect to the camera optical axis are recorded on the pixel which is far from the optical axis of the individual microlenses. Therefore, the light ray traveling along the optical axis of the camera is recorded on the pixel which is exactly at the center of the individual microlens (pinhole model, see Chapter 3, Figure 3.4).

The MLA parameters identified in Chapter 4 are used in the image transformation process. Each microlens records a portion of the scene with different views, hence an LF image recorded on a photosensor has different perspective views of the same scene given by Equation 7.1, where  $(u, v)$  = microlens plane,  $(x, y)$  = photosensor plane, and  $(m, n)$  = size of the photosensor. Equation 7.1 can be rewritten by neglecting  $(u, v)$ , which represents intensity information on the 2D photosensor given by in Equation 7.2.

$$L_F = L(u, v, x, y) \quad \forall (x, y) \in (m, n) \quad 7.1$$

$$L_F = L(x, y) \quad \forall (x, y) \in (m, n) \quad 7.2$$

The MLA provides the option of changing the position of pixels in the LF image with respect to microlens location or direction of light rays depending on user choice. Within an LF image, pixel arrangements with respect to microlens position generate a constraint within itself that forces the final image to have perspective data which are grouped according to the number of microlens units,  $(\bar{m}_x, \bar{m}_y)$ , in the MLA and the number of pixels under each microlens unit,  $(P_x, P_y)$ . The pixel of each microlens is arranged corresponding to the shape of the MLA unit. In the case of a Lytro LF camera, the microlenses are hexagonal in shape and hence to simplify the burden of addressing all the pixels, the microlens images are cropped into a square image with

help of  $r$ , i.e. the radius of the microlens image is defined by the user or calculated manually .

Once the LF data undergoes cropping, there is slightly reduced pixel information under each MLA unit. Hence  $(m', n')$  gives the new count of total LF data (pixels) instead of  $(m, n)$  which is represented by  $(x', y')$  alternative to  $(x, y)$  in Equation 7.3 . The resulting LF image is accessible using microlens location called the microlens addressing mode,  $(L_{mm})$ . This mode of representing the LF image is given by Equation 7.4, where  $n_i$  defines the number of lenslets corresponding to  $I_c$ . In order to access a particular microlens data under  $L_{mm}$ , the MLA cord  $(I_c)$  details have to be specified to determine the principal axis data in terms of pixels.

$$L_F(x', y') = L(x, y) \quad \forall \begin{cases} (x, y) \leq (m', n') \\ 0 > (m', n') \end{cases} \quad 7.3$$

$$L_{mm}(x', y') = L_F(I_{cx} + P_{x'}, I_{cy} + P_{y'}) \quad \forall \begin{cases} I_{cx} = 1 \dots n_i \\ I_{cy} = 1 \dots n_i \\ P_{x'} = -p_x/2 \dots p_x/2 \\ P_{y'} = -p_y/2 \dots p_y/2 \end{cases} \quad 7.4$$

This LF data is further modified to generate 2D camera views where individual perspective pixels are bundled in a group representing a single camera view. The pixels are arranged with respect to the direction of incoming light rays and the corresponding method is known as direction addressing mode  $(L_{dm})$ , represented by Equation 7.5, where  $n$  represents the half value of the lenslet radius defined by the user corresponding to  $I_c$ .

$$L_{dm}(x', y') = L_F(I_{cx} + P_x, I_{cy} + P_y) \quad \forall \begin{cases} p_x = -n \dots 0 \dots n \\ p_y = -n \dots 0 \dots n \\ I_{cy} = 1 \dots n \\ I_{cx} = 1 \dots n \end{cases} \quad 7.5$$

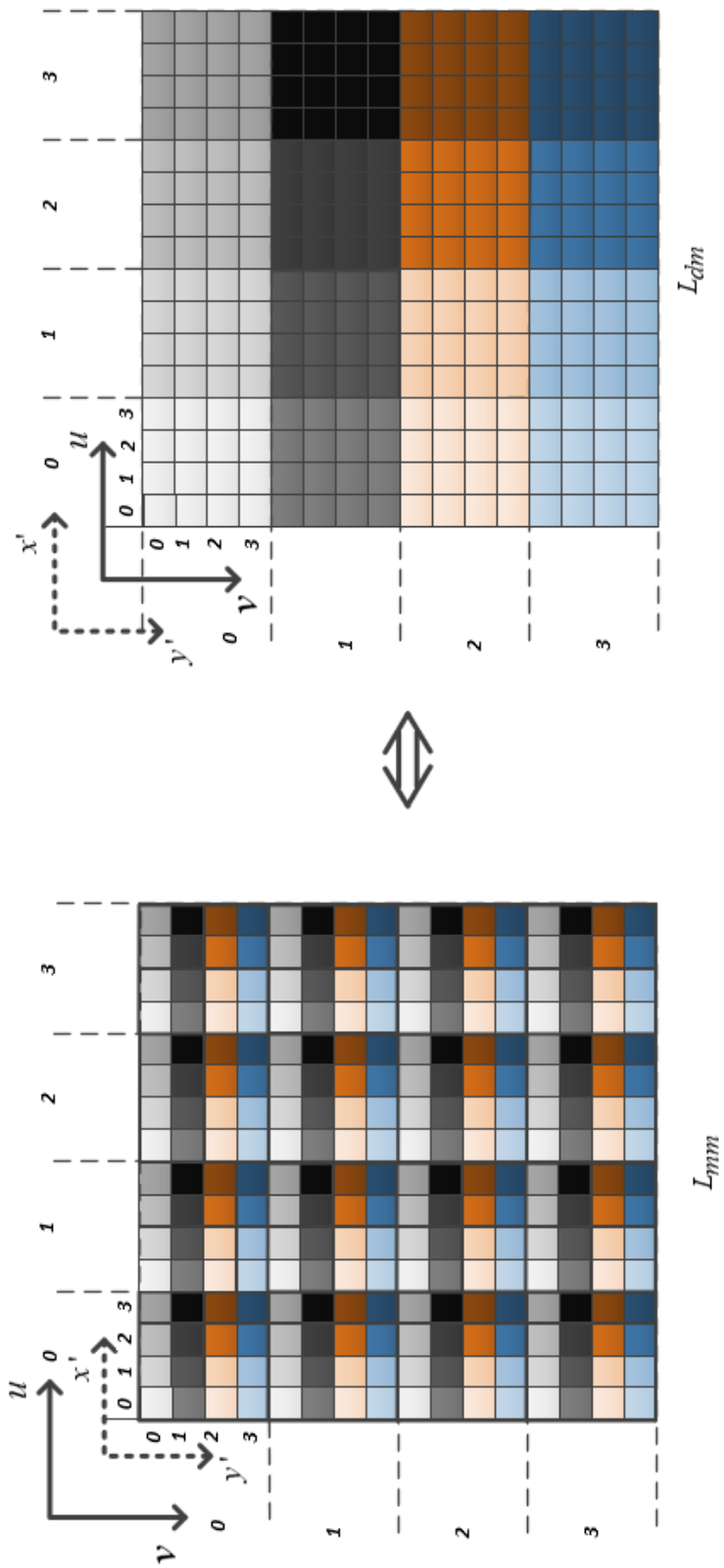


Figure 7.5 Light field data and sub-aperture camera views

In Figure 7.5, the transformed LF data is shown with two possible addressing modes. Every square represents a pixel in the photosensor and a group of these pixels represents the microimages,  $L_{mm}$ . Different views in a particular microlens image are indicated with varying colours which are transformed to form a 2D camera view  $L_{dm}$ . Since the size of the microlens image is 3x3 with 3x3 microimages, the resulting 2D camera view is of 3x3 pixels with 9 camera views. This shows, that the final pixel count in both the addressing modes is not altered, but a representation of directional pixels are changed to generate 2D camera views.

The LF image obtained using Equation 7.5 consists of multiple images, each representing data of specific direction. Hence the entire image can be split into individual camera views that can be used for image processing applications. The lens distortion generated by the main lens is uniformly spread across every individual image generated and hence calibration of each camera view with respect to the main lens is necessary. Let  $(x, y)$  be the pixels under consideration for a particular direction  $\delta$  for a given microlens. Using Equation 7.6 on the entire image results in free floating images which are independent of LF constraints and represents unique relationships with the world scene. Equation 7.7 represents such free floating images sorted with respect to the direction of light rays and Equation 7.8 shows an example to generate one such image.

$$L_{dm}(x_{p_{x+\delta}}, y_{p_{y+\delta}}) = L_F(I_{cx} + P_{x'}, I_{cy} + P_{y'}) \forall \begin{cases} I_{cx} = 1 \dots n_i \\ I_{cy} = 1 \dots n_i \\ P_{x'} = 0 \dots p_x \\ P_{y'} = 0 \dots p_y \end{cases} \quad 7.6$$

$$L_{dm}(x_{p_{x+\delta}}, y_{p_{y+\delta}}) = \begin{bmatrix} L_{dm}(x_1, y_1) & \dots & \dots & L_{dm}(x_1, y_{p_{y+\delta}}) \\ \vdots & \vdots & \vdots & \vdots \\ L_{dm}(x_{p_{x+\delta}}, y_1) & \dots & \dots & L_{dm}(x_{p_{x+\delta}}, y_{p_{y+\delta}}) \end{bmatrix} \quad 7.7$$

$$L_{dm}(x_{p_{x+\delta}}, y_{p_{y+\delta}}) = L_{dm}(1_1, 1_1) \forall \left\{ \begin{array}{l} (x, y = 1) \\ (p_{x+\delta}, p_{y+\delta} = 1) \end{array} \right\} \quad 7.8$$

A free-floating camera view (image) of an LF data is very important in this research because each of these camera views will have all optical properties as of the original raw LF image and can be represented without using specific variables (microlens location, light direction) further to this step.

#### 7.4. Calculating initial calibration parameters – Step 1

The classification of LF images according to the direction of light rays/perspective views provides individual camera view ( $\varphi_i$ ) images which are independent of the microlens arrangement. All camera view images generated from LF images will equally share the optical property (distortion) that was initially part of the complete LF image and hence calibration has to be applied to all camera views. All camera views will have specific calibration parameters which are unique to a specific direction of light rays. Every camera view ( $\varphi_i$ ) is considered as a single 2D image and a group of camera views of a LF camera is represented by Equation 7.9.

$$\varphi_{1\dots i} = \begin{bmatrix} \vdots & \dots & \vdots \\ (\varphi_{-x}, \varphi_{-y}) & (\varphi_0, \varphi_0) & (\varphi_x, \varphi_y) \\ \vdots & \dots & \vdots \end{bmatrix} = L_{dm}(x_{p_{x+\delta}}, y_{p_{y+\delta}}) \quad 7.9$$

With the transformation of the LF image into 2D camera views the microlens effect on the LF camera system will be changed, i.e before transformation the image size of each individual microlens was  $(p_x, p_y)$  but due to the transformation, that image size will be increased to a size given by Equation 7.10. In this method the physical lens is not replaced after transformation, hence the term virtual lens can be used without affecting the original camera elements (lenses).

$$\varphi(u,v) = \left( m'/p_x, n'/p_y \right) \quad 7.10$$

With the introduction of the virtual lens replacing the MLA unit, this helps the LF camera system to be considered as a two lens system, where each virtual lens system acts as directional filters allowing the light rays of a particular direction (reflected angle) to be recorded on the photosensor. The light rays recorded using a virtual lens are independent of each other and hence while calculating calibration parameters, directional information for each 2D image can be neglected.

The directional information can be extracted by comparing with a standard reference, but to increase the efficiency and save computational time it is recommended to store directional information while transforming LF images. The transformed LF images of a checkerboard, which were captured using the LF camera are compared with calibrated checkerboard to obtain the orientation of free-floating images

#### 7.4.1. Double Pinhole Model

The LF camera system consists of two lens elements in the system (main lens and MLA) which bend the incoming light rays towards the optical centre of the individual lens system. Due to the transformation applied (Sec 7.3.2) to convert the LF image into specific 2D camera views, and with the introduction of a virtual lens that will allow only specific directional light rays onto the photosensor, the arrangement resembles a two-lens system where the second lens receives directional light rays from the main lens to represent 2D camera images [131].

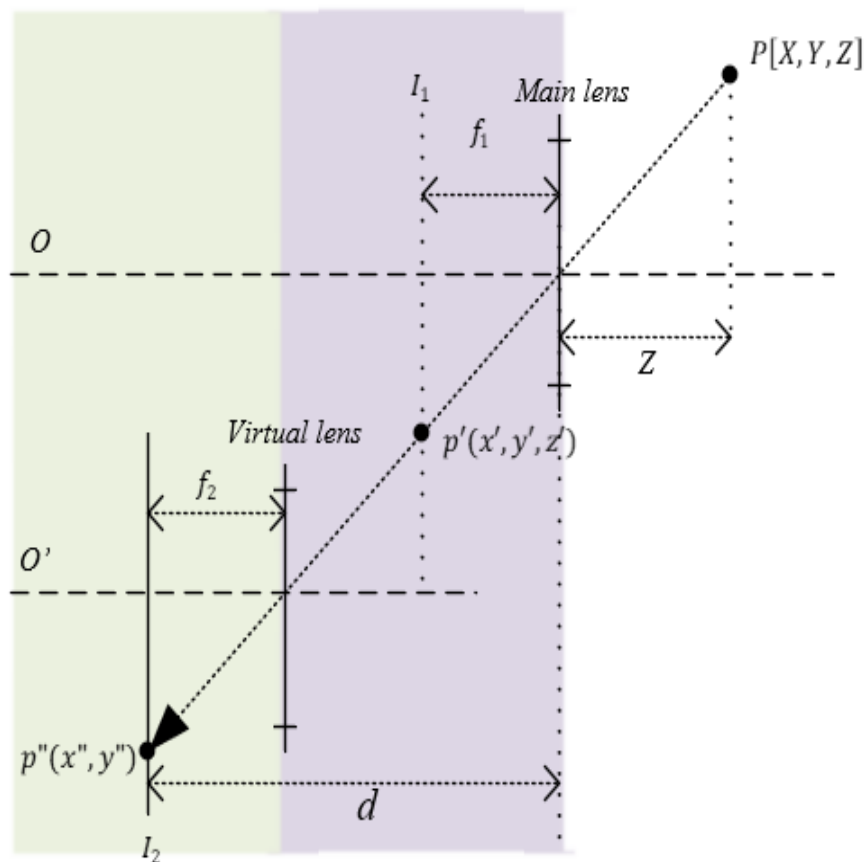


Figure 7.6 Double pinhole model



Consider, if a world point object  $P[X, Y, Z]$  at distance of  $Z$  from the main lens (pinhole) is imaged on the camera's image plane ( $I_1$ ) at the coordinate  $p(x', y')$  it would represent a normal 2D camera system. The camera parameters that map 3D features on to a 2D plane can be found using similar triangles given by Equation 7.11. Using homogeneous coordinates for  $p(x, y)$ , Equation 7.11 is transformed to Equation 7.12 [111, 120].

$$x' = \left(\frac{Xf_1}{Z}\right) \quad y' = \left(\frac{Yf_1}{Z}\right) \quad 7.11$$

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} f_1 & 0 & 0 \\ 0 & f_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad 7.12$$

If the light ray is not imaged on the image plane and the medium inside the camera space is air, then the light ray would be represented with world coordinates in the camera space, represented by  $p'(x', y', z')$ . Since the light ray has only interacted with the main pinhole, the depth details can be collected from the light rays and hence  $z' = f_1$  at the image plane  $I_1$ . The virtual lens can be considered acting as a pinhole model is placed at a distance of  $z''$  from the image plane  $I_1$  receiving the same directional light ray as shown in Figure 7.6. With the help of the photosensor at the image plane  $I_2$ , the light ray is captured at a distance of  $f_2$  from the vitural lens system. Let  $p''(x'', y'')$  be the pixel where the light rays are recorded which represents a 2D coordinate with respect to camera space. Using similar trianlge rules,  $x''$  and  $y''$  are given by Equation 7.13 where  $d$  is the distance between the main lens and the MLA. Equation 7.13 represents values of  $p''(x'', y'')$  with respect to the world object  $P[X, Y, Z]$ .

$$x'' = \left(\frac{Xd}{Z}\right) \quad y'' = \left(\frac{Yd}{Z}\right) \quad 7.13$$

Hence  $p''(x'', y'')$  needs to be translated to the desired origin. Let this translation be defined by  $(t_u, t_v)$ , given by Equation 7.14.

$$x'' = X\left(\frac{d}{Z}\right) + t_u \quad y'' = Y\left(\frac{d}{Z}\right) + t_v \quad 7.14$$

The pixels of the camera is indicated as pixels/mm. If the pixels are square the resolution will be identical in both directions of the camera image coordinates. However, for a more general case, rectangle pixels are considered with a resolution  $m'_u = m'$  and  $m'_v = n'$  pixels/mm. Hence Equation 7.14 will be updated with resolution data given by Equation 7.15. The homogenous equations of  $p''$  similarly to  $p'$  is given by Equation 7.16.

$$x'' = m'_u X \left( \frac{d}{Z} \right) + m'_u t_u \quad y'' = m'_v Y \left( \frac{d}{Z} \right) + m'_v t_v \quad 7.15$$

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} m'_u d & 0 & m'_u t_u \\ 0 & m'_v d & m'_v t_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} R = KR \quad 7.16$$

Note that  $K$  only depends on the intrinsic camera parameters such as focal length, and principal axis, and thus defines the intrinsic parameters of the camera. Sometimes  $K$  also has a skew parameter ( $S$ ) (shear distortion in the projected image [126]), given by Equation 7.17.

$$\mathbf{K} = \begin{bmatrix} \alpha_u & S & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad 7.17$$

To calculate the external parameters of the camera, the method suggested by Zhang [121] has been followed by a 3x4 matrix represented by  $E = (R|RT)$ . The final camera transformation can now be represented by  $\mathbf{K}(R|RT) = (\mathbf{K}R|\mathbf{K}RT) = \mathbf{K}R(I|T)$ , given by Equation 7.18 where  $C$  is a 3x4 matrix having 12 entries, the entry of 3<sup>rd</sup> row and 4<sup>th</sup> column is 1, thereby 11 unknown parameters.

$$\bar{P} = \mathbf{K}R(I|T) = CP \quad 7.18$$

Since the 2D camera view is a part of the total LF image, the lens distortion of the main lens and the MLA lens units are equally distributed among all the 2D camera views. So the distortion correction is for all of the 2D camera views individually and saved according to the light rays directional data saved earlier.

The distorted points are denoted as  $(x_{distorted}, y_{distorted})$  where  $x, y$  are undistorted pixel locations and they are in normalized image coordinates. Normalized image coordinates are calculated from pixel coordinates by translating to the optical center and dividing by the focal length in pixels, thus  $x$  and  $y$  are dimensionless.  $k1, k2$ , and  $k3$  are the radial distortion coefficients (Equation 7.19) of the lens and  $p1, p2$  are the tangential coefficients (Equation 7.20), while  $r^2 = x^2 + y^2$ .

$$x_{distorted} = x(1 + k1r^2 + k2r^4 + k3r^6) \quad 7.19$$

$$y_{distorted} = y(1 + k1r^2 + k2r^4 + k3r^6)$$

$$x_{distorted} = x[2p1xy + p2(r^2 + 2x^2)] \quad 7.20$$

$$y_{distorted} = y[p1(r^2 + 2y^2) + 2p2xy]$$

The resulting image after distortion correction represents the world point  $P$  imaged on the photosensor behind the virtual lens. Equation 7.18 represents the combined focal length of both lenses, the main lens and the virtual lens (MLA). Due to the 2D transformation of LF data, this results in a two pin hole system, hence calibration of each 2D view is necessary which results in a large stack of calibration parameters  $\bar{p}_{total}$  as given by Equation 7.21, where  $i, j$  represents the number of 2D camera views which is equal to the number of pixels under each microlens unit  $(p_x, p_y)$  after cropping.

$$\bar{p}_{total} = \begin{bmatrix} \bar{p}_{(1,1)} & \cdots & \bar{p}_{(1,j)} \\ \vdots & \ddots & \vdots \\ \bar{p}_{(i,1)} & \cdots & \bar{p}_{(i,j)} \end{bmatrix} \quad 7.21$$

## 7.5. Multi stereo lens calibration - Step 2

The calibration parameters of LF data found in the previous step are applied to all the camera views. In the second step, with the help of directional data, a particular 2D camera view and the corresponding calibration parameter is chosen from the stack of data as shown in Equation 7.22.

$$\varphi'_{(i,j)} = \sum_{i=p_x+\delta} \sum_{j=p_y+\delta} \bar{p}_{(i,j)}(\varphi) \quad 7.22$$

Where  $i = p_{x+\delta}$  and  $j = p_{y+\delta}$  represent the direction of light,  $\bar{p}_{(i,j)}$  represents the calibration parameter selected based on directional data  $i, j$  and  $\varphi$  is the selected camera view, while a distortion-free image of a particular view as represented by  $\varphi'_{(i,j)}$ .

### 7.5.1. Preferences of input images

In step-1 of calibration, the LF image was calibrated for lens distortions that may have been associated with the camera, due to lens defect or due to mounting issues (see Chapter 4, Sec 4.4 for mounting issues). After processing images with step-1 calibration parameters, individual cameras views ( $\varphi'$ ) are well calibrated with world reference, i.e. the straight line looks approximately straight in the raw data. From this stage onwards, camera view images are referred to as sub-aperture images, because the location and direction parameters play an essential role in the calibration step-2.

These calibrated sub-aperture images are used for post-focusing techniques, where light rays (represented by pixels) of a camera view interact with the neighbouring camera views to generate a virtual image at a user-defined distance/focal length. Hence sub-aperture images must be calibrated with reference to neighbouring sub-aperture images. This means the raw sub-apertures images must be compared with each other to understand their alignment. The central view is considered as a primary reference image ( $PR$ ), among all the sub-aperture images because light rays from two pinhole lenses form a central view image. The sub-aperture images that align with the central view (in the same row and column direction) are considered as second reference images ( $SR$ ). The result generated by comparing  $PR$  and  $SR$  images provides necessary base disparity information in terms of pixels, that play an important role in calibrating the raw image with respect to the MLA physical parameters. While remaining camera views are referred to as non-reference images ( $TR$ ). The base disparity information varies depending upon the distance between the MLA and the photosensor. A rigid construction of MLA-photosensor assembly exhibits fixed base disparity (see Chapter 6) that remains constant for a given camera. The sub-aperture images should produce disparity values close to the base disparity value to ensure a good calibration of the LF image.

In a practical scenario, the base disparity values are not provided by the LF cameras manufacturers and hence it becomes important to calculate manually. The input images used for calibration step-2 must be taken with targets (checkerboard patterns) parallel to the camera. This will ensure the disparity values recorded by cameras are for the same depth in the Z direction. If the LF cameras use a fixed lens system, the base disparity calculation is a single step procedure and requires 5 to 10 image sets. In contrast, for the LF camera using a CLS, such as the Lytro family of cameras, the number of images required is higher (15 to 20) taken at a different distance from the target with TTF feature activated manually while capturing a target image. Figure 7.7 represents the distribution of light rays when the target plane is at a slanted angle with respect to the camera plane. It can be noticed a varying distribution of white and black pixels within every microimage of far and near enlarged views (Figure 7.7, bottom). For calculating base disparity information, the target distance from camera plane is flexible, while the orientation of the target plane must be parallel to obtain a uniform distribution of light rays (white and black pixels, in this case) across the target surface considered for calibration.

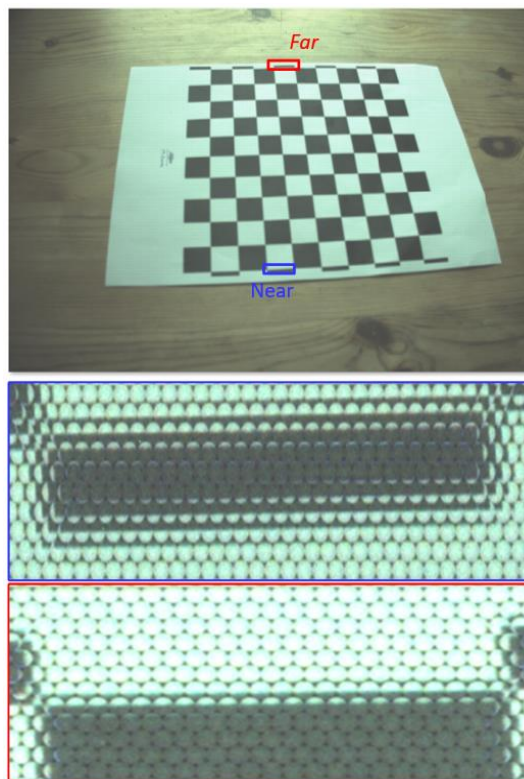


Figure 7.7 Near and far regions of raw LF image representing the distribution of light rays when targets are a slant to camera plan

In a practical scenario, aligning the target plane parallel to the camera plane is difficult. Hence, the target is manually aligned parallel to the camera plane and correct alignment is made with the help of the homography ( $h$ ) method between sub-aperture image and synthetically generated image (see Appendix 10.6). A synthetic image is generated by user-defined grid sizes corresponding to the grid size used by the targets. The user selected control points in the sub-aperture image and corresponding synthetic data are matched to generate a matrix ( $h$ ), that transforms the sub-aperture image parallel to camera plane. In Figure 7.8, one of the Illum sub-aperture image and synthetic data are shown with control points highlighted with red circles (left and middle). Using the homography method on the input image with base accuracy (see Chapter 6) of 0.4345 the image has been transformed into an image with a base accuracy of 0.5715. Here, the input image used was not calibrated with step-1 parameters representing the worst case, but in the practical scenario, the input images are always calibrated with step-1 and hence base accuracy is high approximately greater than 0.7 with 1.0 being the maximum score. The homography method is usually recommended for calibrating camera for general applications, while for engineering applications input images should have base accuracy greater than 0.75, this value is being defined with reference to the data shown in Figure 7.9 without applying the homography method.

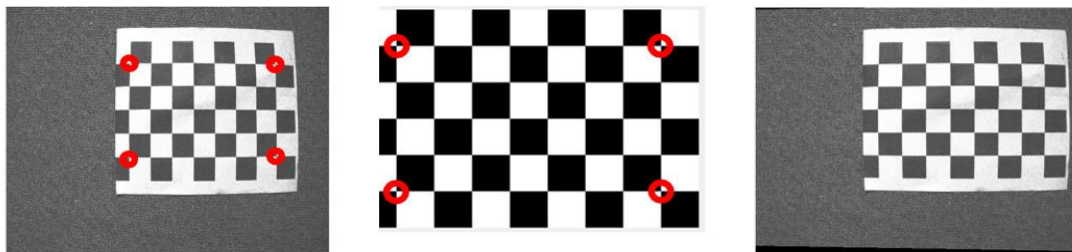


Figure 7.8 Representing control points in a sub-aperture and corresponding synthetic data to generate target plane parallel to the camera plane

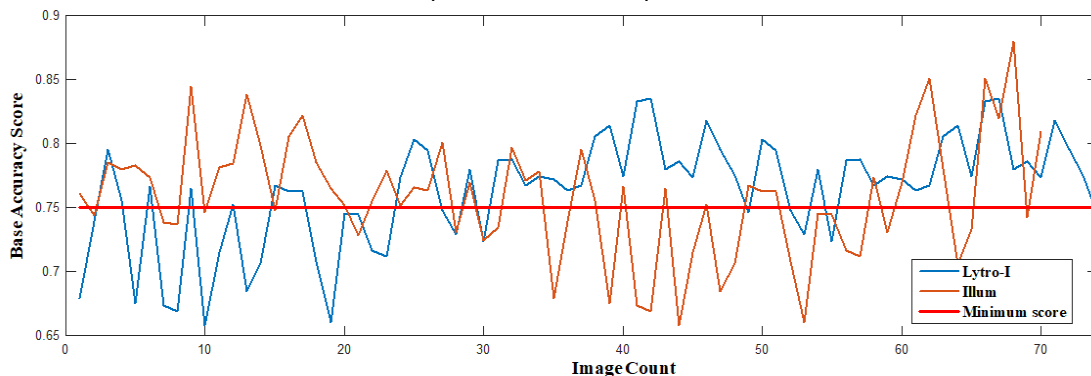


Figure 7.9 Representing input image base accuracy score for the Lytro family of cameras

As mentioned earlier, all *SR* images are compared with *PR* images according to their respective locations on the  $(u, v)$  plane. In a LF image, the pixel count of the microimage corresponds to the number of neighbouring (sub-aperture) images with respect to the *PR* image. The microimage resolution of the Lytro-I and the Illum cameras are  $7 \times 7$  pixels and  $13 \times 13$  pixels respectively. Hence consisting of 3 and 6 neighbouring images groups around the *PR* image known as Neighbourhood (*NH*), as shown in Figure 7.10. A group of square boxes with similar colour, surrounding *PR* in Figure 7.10 (top) represent a *NH* group. The classification of the sub-aperture image with respect to *PR* is represented by Figure 7.10 (bottom left) and Figure 7.10 (bottom right) for the Illum and the Lytro-I generation cameras respectively. It can be observed that the Illum result consists of  $7NH$ , while the Lytro-I generation result has  $3NH$ .

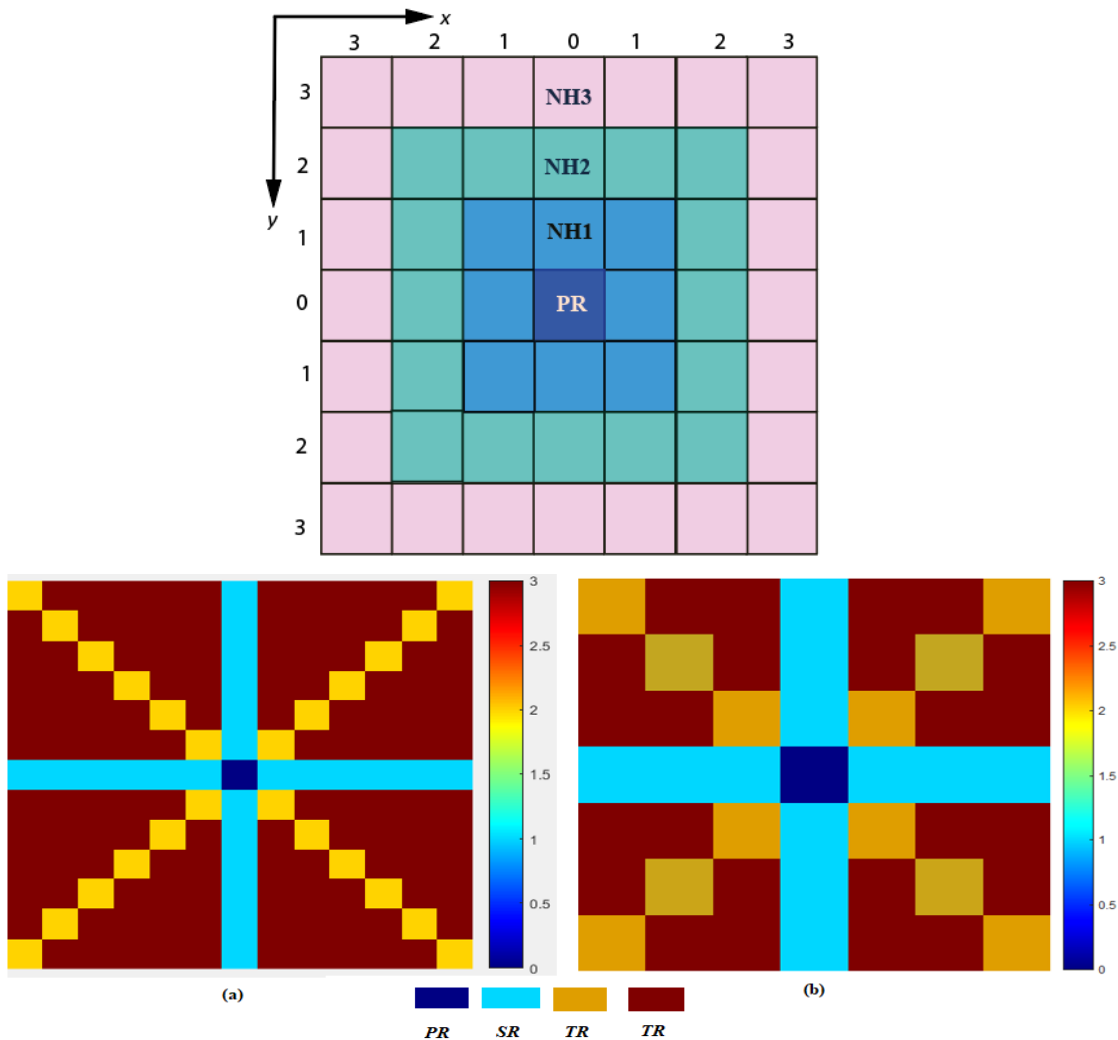


Figure 7.10 Represents a group of Neighbourhood (top), *PR*, *SR* and *TR* for Illum (a) and Lytro-I generation (b) cameras

### 7.5.2. Base disparity

The base disparity term was introduced in Chapter 6 to develop the relationship between different perspective images that were generated using the LDS. In this chapter, since the raw image and the corresponding MLA parameters are known, generating a relationship between sub-aperture images is straightforward and has been identified while transforming the LF images. The sub-aperture (perspective) images exhibit a constant shift when *PR* images are overlapped with *SR* and *TR* images to match the feature points. The checkerboard corner points are considered as features and the shift generated is known as base disparity ( $b_d$ ). The base disparity values are always minimum at the focus plane and varies on either side of the focus plane selected using the TTF (see chapter 6).

To calculate the base disparity only *PR* and *SR* images are considered. Since the MLA lenslets are rigidly packed with a constant distance between lenslets [73], the constructional feature of the MLA is considered as a reference to determine the base disparity. A schematic representation of *PR* and *SR* images is shown in Figure 7.11, where only two neighbourhood sub-aperture views surrounding *PR* are considered for simplicity. A cartesian coordinate system ( $x, y$ ) is used to represent various *SR* images. Due to the constructional similarity of the lenslets, the distance measured between *NH1* (images surrounding *PR* immediately) and *PR* images, on all axes, must be equal. To increase the accuracy of measurement, the feature points generated in an image are matched with *PR* and immediate neighbouring images feature points, i.e. the *SR1* image is at a distance of  $b_d$  from the *PR*, while the same distance must show up upon matching with *SR2* and *SR3*. When comparing images of *NH* value greater than 2, the base disparity values must be normalised using a distance metric, given by Equation 7.23 where *NH* represents neighbourhood value.

$$b_d = \frac{\text{feature points}}{NH} \quad 7.23$$



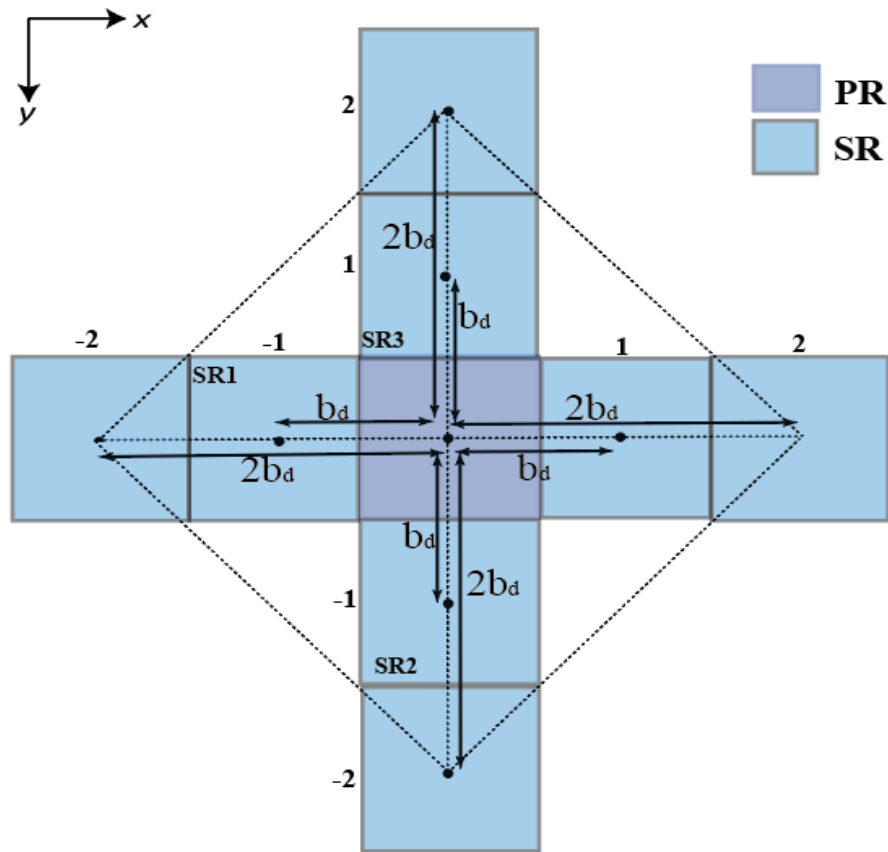


Figure 7.11 Base disparity estimation sketch

---

### Algorithm for calculating base disparity

---

- 1: Load image sets ( $\varphi'_{xy}$ )
  - 2: set  $b_d \Rightarrow 0.75$
  - 3: for  $i=1: N$
  - 4: find  $P_0$  and  $P_{1\dots n}$
  - 5: calculate disparity ( $d_x, d_y$ ) between  $P_0$  and  $P_{1\dots n}$
  - 6:  $(y_{NH}^+) = \varphi'_{xy}(P_0 - P_{1\dots n})/NH$ , if  $(c \geq \varphi'_y)$
  - 7:  $(y_{NH}^-) = \varphi'_{xy}(P_0 - P_{1\dots n})/NH$ , if  $(c \leq \varphi'_y)$
  - 8:  $(x_{NH}^+) = \varphi'_{xy}(P_0 - P_{1\dots n})/NH$ , if  $(c \geq \varphi'_x)$
  - 9:  $(x_{NH}^-) = \varphi'_{xy}(P_0 - P_{1\dots n})/NH$ , if  $(c \leq \varphi'_x)$
  - 10: endfor
  - 11:  $b_d = \text{mean}(y_+, y_-, x_+, x_-)$  by removing outlier
- 

Figure 7.12 Algorithm for base disparity estimation

For calculating  $b_d$ , all images that are calibrated in step1 ( $\varphi'$ ) were processed to generate base disparity values ( $b_a$ ) with a threshold value of 0.75 or greater without using homography methods stated earlier (line 2, Figure 7.12). The checkerboard corner data  $P_0$  forms  $PR$  image, and  $P_{1\dots n}$  for the rest of the  $SR$  images are generated using method [111]. Since the perspective views were shifted by a few pixels compared to the central view, some of the corner points present in  $P_0$  did not match with  $P_1$  due to noise. Hence, every corner point in  $P_0$  were compared with  $P_1$  to find the potential match with minimum Euclidean distance using Equation 7.24 and Equation 7.25.

$$P_1(x, y) = \underset{\min}{\operatorname{arg}} \left| \sqrt{(P_{0x} - P_{1\dots nx})^2 + (P_{0y} - P_{1\dots ny})^2} \right| \quad 7.24$$

$$P_1(x) \leftrightarrow P_{0x} \quad 7.25$$

$$P_1(y) \leftrightarrow P_{0y}$$

The values  $d_x$  and  $d_y$  were calculated by subtracting the lower value from the higher feature point value. This ensures that, when disparity values are plotted, the points must stay at the positive axis to exhibit non-zero crossing feature. The calibration image set was forced manually to exhibit a single view as explained in Sec 7.3.2. Using the conditions provided in line 6 to 9 of Figure 7.12, the disparity values are grouped with respect to the parent camera view location with respect to the  $PR$  image, i.e.  $y_{NH}^+$ ,  $y_{NH}^-$ ,  $x_{NH}^+$  and  $x_{NH}^-$  where  $NH$  represents the neighbourhood value. Upon grouping, due to the image processing steps involved while calculating checkerboard corners, noise features, and outliers, are introduced into feature points generated. Hence a robust linear fit model was applied to negate the outliers and consider the strong and recurring points. An initial threshold distance of 0.0 was deliberately given to make the model robust.

Figure 7.13 and Figure 7.14 represent the disparity score generated using 10 image sets with base accuracy greater than 0.8. Some of the important observations to be made are:

1. Since the majority of the feature points are on the positive axis (positive disparity), the transformation of the raw image into sub-aperture was carried in the right order with respect to the central view.

2. The feature points in the graph exhibit variations, this is due to the image processing techniques applied while transforming the LF data and corner detection. But the variations are approximately  $\pm 0.2$  pixels, with some feature points having greater variations and they are regarded as outliers. A voting system was adapted to choose a consistent number from the feature point disparity space. This issue can be solved by using a large image set with better base accuracy.
3. Since disparity value depends on the parent camera view location and  $NH$  value, it can be observed that the disparity values increase as the  $NH$  increases (for example, denoted as  $Y1, Y2$  to  $Y6$ ). This supports the main concept of multi stereo calibration, that disparity values increase with constant rate towards the outermost  $NH$  image.

Since a large stack of feature data was generated during the process, only the first 800 features are shown in the graphs to illustrate noise and variations in the feature disparity space. Since the Illum camera has a  $NH6$  value, only the  $Y +$  axis is shown in Figure 7.13 (excluding first  $NH$ ), while all  $NH3$  images are shown for the Lytro -I generation cameras.

Figure 7.15 and Figure 7.16 represent the final results of the disparity estimation algorithm over different  $NH$  values. It can be observed in Figure 7.15 for the Illum camera that, the disparity values rise steadily from around 0.46 pixels  $NH1$  and continue linearly with minor deviation. The expected line is drawn to indicate the constant rate of disparity (theoretical values) by fitting a straight line to first  $NH3$  disparity values. The graph suggests that, Illum camera view can be considered for further use without calibration step-2, but to use camera views after the 4<sup>th</sup>  $NH$ . In contrast, the Lytro-I generation cameras exhibit efficient correlation with surrounding camera views as indicated by the steady disparity line. Since, the Lytro-I has got a maximum of  $NH3$  values it appears to generate better disparity after calibration step-1. In comparison, the Illum has good correlation with neighbouring images until  $NH3$  and descends thereafter.

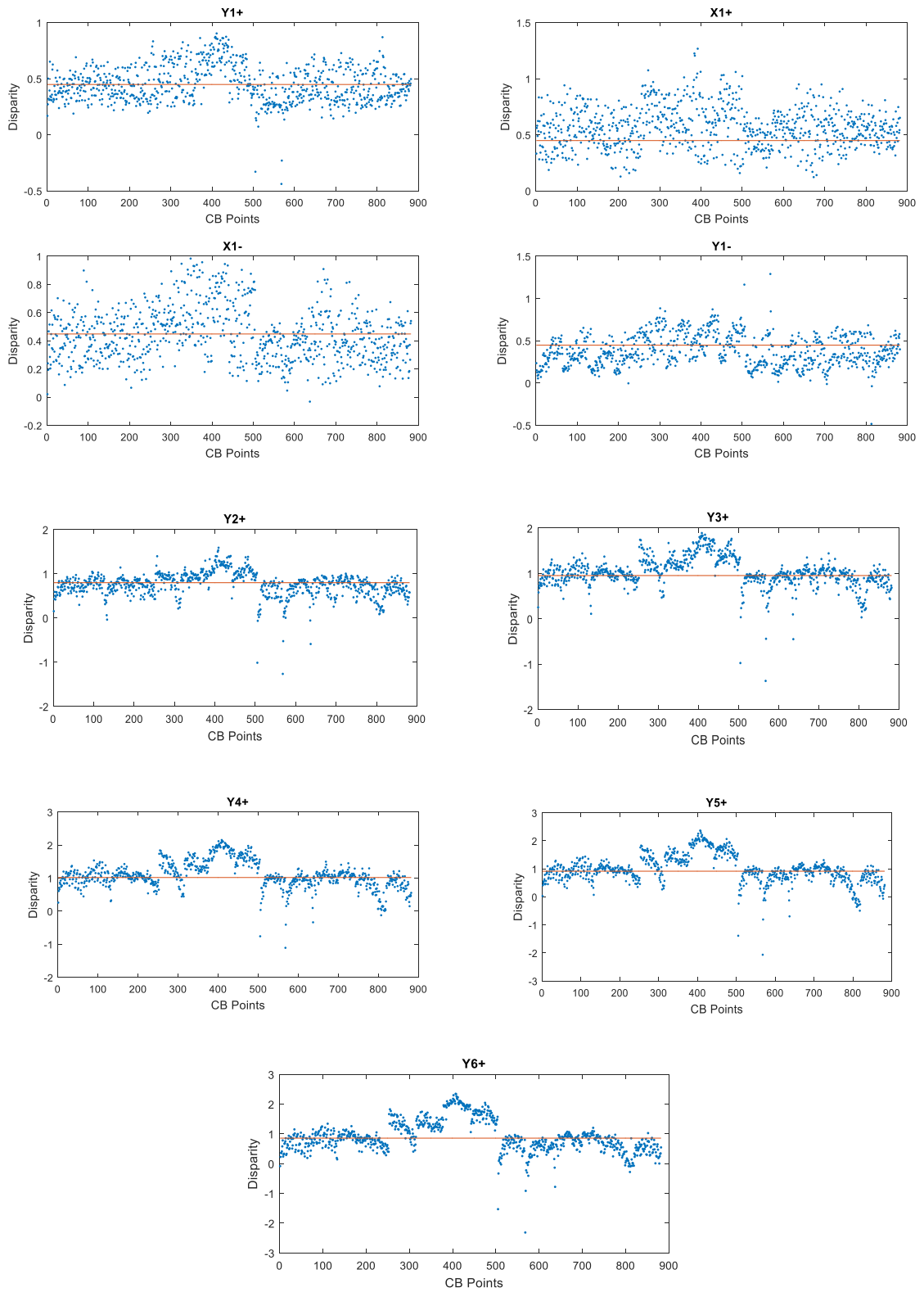


Figure 7.13 Graphs showing the distribution of the feature points in the various neighbourhood images and the corresponding mean line for the Illum camera.

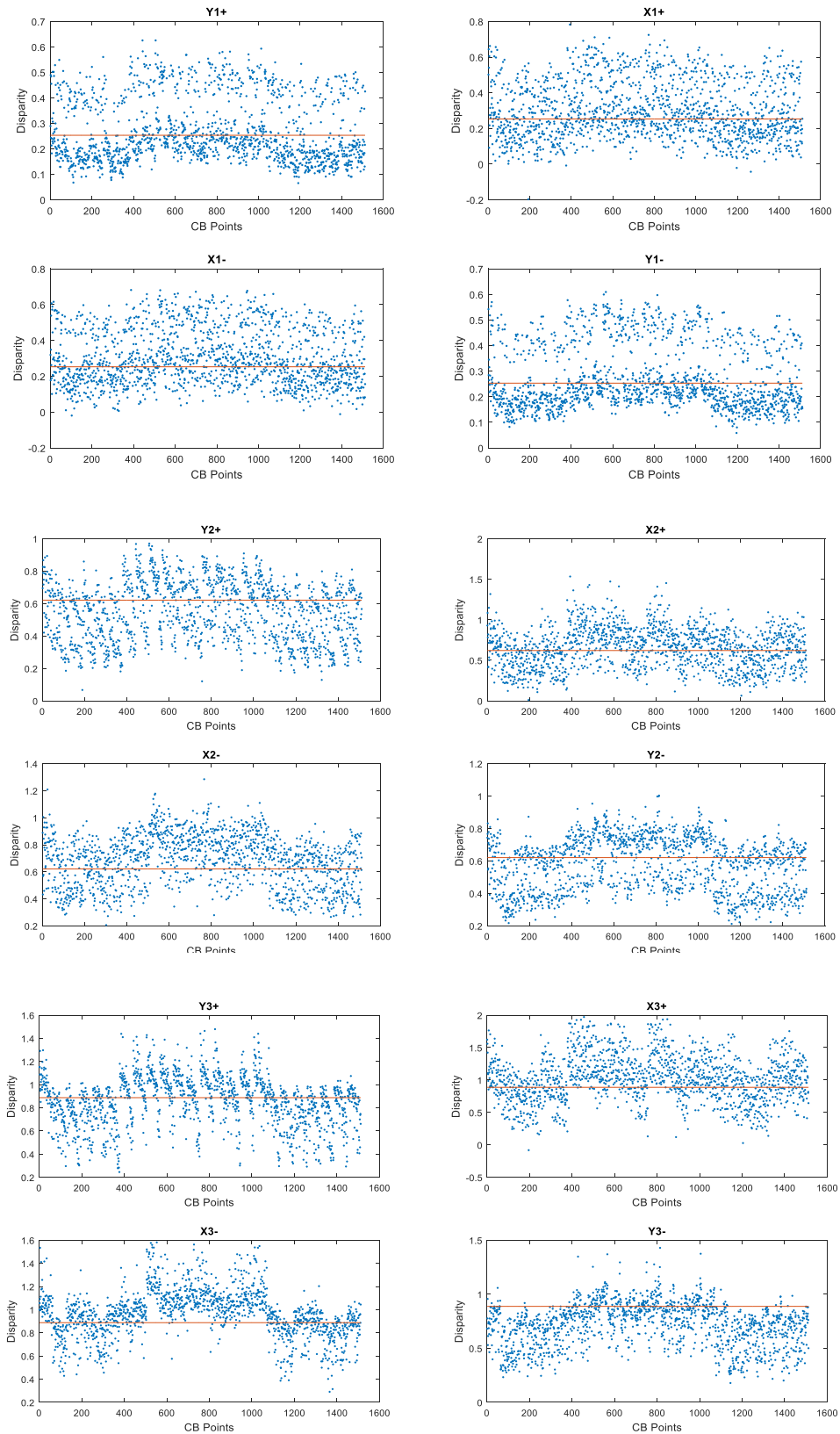


Figure 7.14 Graphs showing the distribution of the feature points in the various neighbourhood images and the corresponding mean line for the Lytro-I camera.

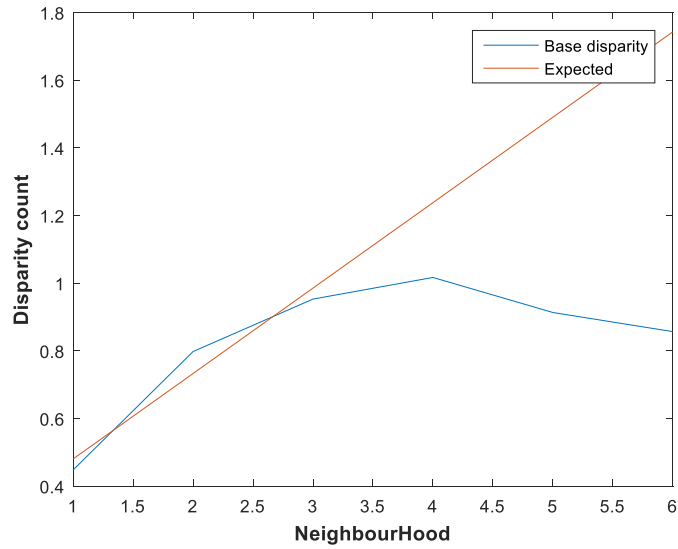


Figure 7.15 Base disparity estimation of different neighbourhood groups for the Illum camera

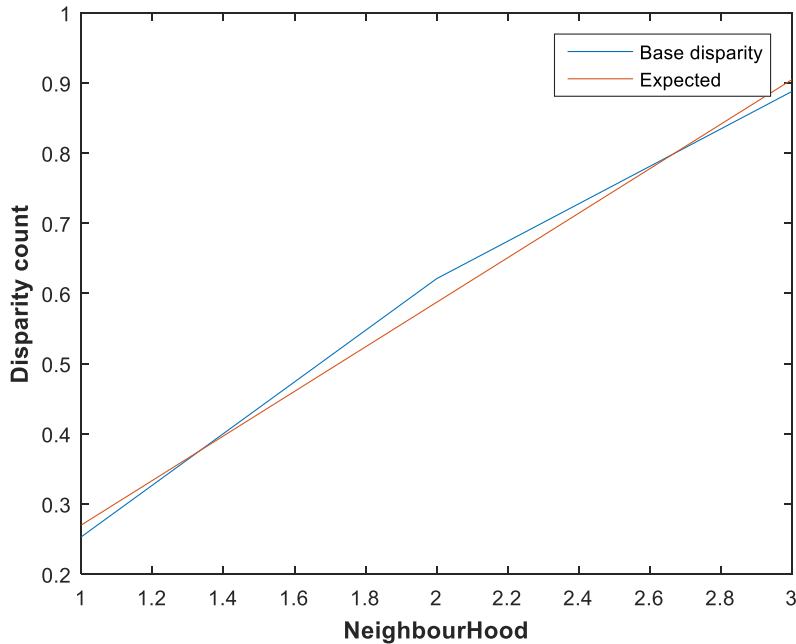


Figure 7.16 Base disparity estimation of different neighbourhood groups for the Lytro-I camera

### 7.5.3. Updating radial distortion coefficients

The radial and tangential parameters obtained in the step-1 will help to calibrate raw images with respect to the world plane  $(X, Y, Z)$ , where individual camera views are considered. By calculating the base disparity, the necessary validation can be generated as shown in Sec.7.5.2, this is because lenses are not perfect pinhole models and hence light rays pass through the lenslets unevenly.

The feature points are generated for all *SR* and *TR* images, representing distorted points (*dp*). By comparing these distorted points with the base disparity values, the expected location of the feature points is calculated, and they represent the distortion-free points (*dfp*). These points can be represented as Equation 7.26 by using Equation 7.19 and Equation 7.20, where  $x_0$  and  $y_0$  represent the principal point.

$$\begin{bmatrix} (dfp_x - dfp_{x0})(r) & dfp_x - dfp_{x0}(r^2) \\ (dfp_y - dfp_{y0})(r) & dfp_y - dfp_{y0}(r^2) \end{bmatrix} \begin{bmatrix} K1 \\ K2 \end{bmatrix} = \begin{bmatrix} dp_x - dfp_x \\ dp_y - dfp_y \end{bmatrix} \quad 7.26$$

This is in the form of  $Ax = b$  and can be solved using least-square solutions for  $K1$  and  $K2$  [126]. Figure 7.17 represents the feature points plotted for distorted points (red) and points generated after using  $K1$  and  $K2$  (Equation 7.26) to generate a distortion-free image. An enlarged view shows the symmetry between calibrated feature points.

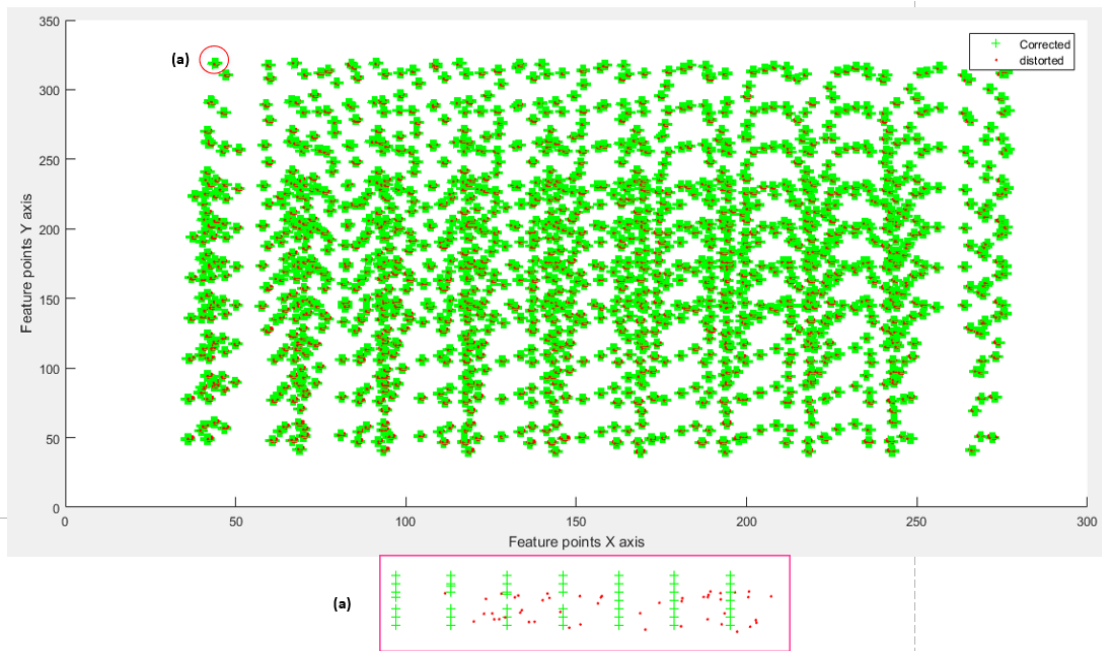


Figure 7.17 Feature points of the calibrated image (step-2)

## 7.6. Results

The fundamental theme of this calibration method is to use the basic parameters of an LF camera (MLA parameters) that remains unaffected for a given photosensor and MLA assembly. The calibration method illustrated in this chapter uses the step-1 calibration method that removes. While in step-2, light rays are forced to follow the pinhole model by comparing the base disparity for a flat surface parallel to the camera. The basic MLA parameters used in the calibration procedure are shown in Table 7 2.

Table 7-2 Basic MLA parameters of the Lytro family used in the calibration method

Camera	MLA Pitch (Pixels)	Microimage (Pixels)	Microlens (Pixels)
Lytro-I	10x10	100	376x327
Illum	13x13	169	434x541

In the calibration method illustrated here, different sizes of white-black square grids (checkerboards) were used to evaluate the root mean square (RMS) ray re-projection error compared with the checkerboard size. Since there are two calibration steps involved in this method, the re-projection error values were calculated using images that are generated after step-2 (undistorted images). For the applications where distortion-free images are required, only the first stage of calibration is necessary, and for the applications where stereo vision and post-focusing are the primary goals then both the calibration, stages are necessary. The mean re-projection error of calibration of 2D camera views is shown in Table 7-3 over different parameters.

The re-projection error evaluated using this method is very close to the results of [66, 82]. But the advantage of this calibration method over other methods is illustrated in Figure 7.19 and Figure 7.20 in comparison with Figure 7.18. The method followed in [85] is restricted to the alignment of the light rays passing through the pinhole camera model, but calibration step-2 illustrated in this chapter considers the parent light rays and corresponding neighbouring light rays using virtual lens techniques and base disparity values.



Table 7-3 Calibration parameters generated by our algorithm for the Lytro cameras

Lytro Model	Grid (mm)	Re-projection error	Standard deviation	CPU time (min)
Lytro-I	19.6	0.3132	0.0017	32.8
	35.1	0.3795	0.0079	35.5
Illum	19.6	0.3150	0.0497	62.5
	35.1	0.3163	0.0302	64.6

Figure 7.18 represents the Lytro-I re-projection error of all sub-aperture views generated using line features, calculated using the method described in [85]. It can be observed that the re-projection error increases over the distance from the central view (0.3 to 0.9 pixels). This behaviour is because light rays tend to deviate moving away from the camera optical centre, and hence euclidean distance increases between the expected and actual trajectory. This can be minimised by using calibration step-2 illustrated here and the re-projection error for the Lytro family of cameras is represented in Figure 7.19 and Figure 7.20 where the overall re-projection error between all possible sub-aperture images is approximately  $\pm 0.1$  pixels when compared to 0.9 pixels using [85]. The mean standard deviation (SD) of the re-projection error is represented in Table 7-3.

Table 7-4 base disparity calculated for Lytro family of cameras

Camera	1NH	2NH	3NH	4NH	5NH	6NH
Lytro-1	0.26	0.61	0.78	0.0	0.0	0.0
Illum	0.48	0.94	1.32	1.80	2.12	2.56

A set of 20 checkerboard images were used to calibrate the Lytro family of cameras requiring approximately 64 minutes and 32 minutes for the Illum and Lytro-1 cameras, while it takes 0.12 minutes to apply distortion correction. This novel method is approximately 50% faster compared to [82] using a 2.2 GHz PC with reprojection error of approximately 0.3 pixels. The base disparity calculated for the Lytro family of cameras are tabulated in Table 7-4, where results are in pixels. A zero-pixel disparity indicates the camera has no corresponding neighbourhood images available.

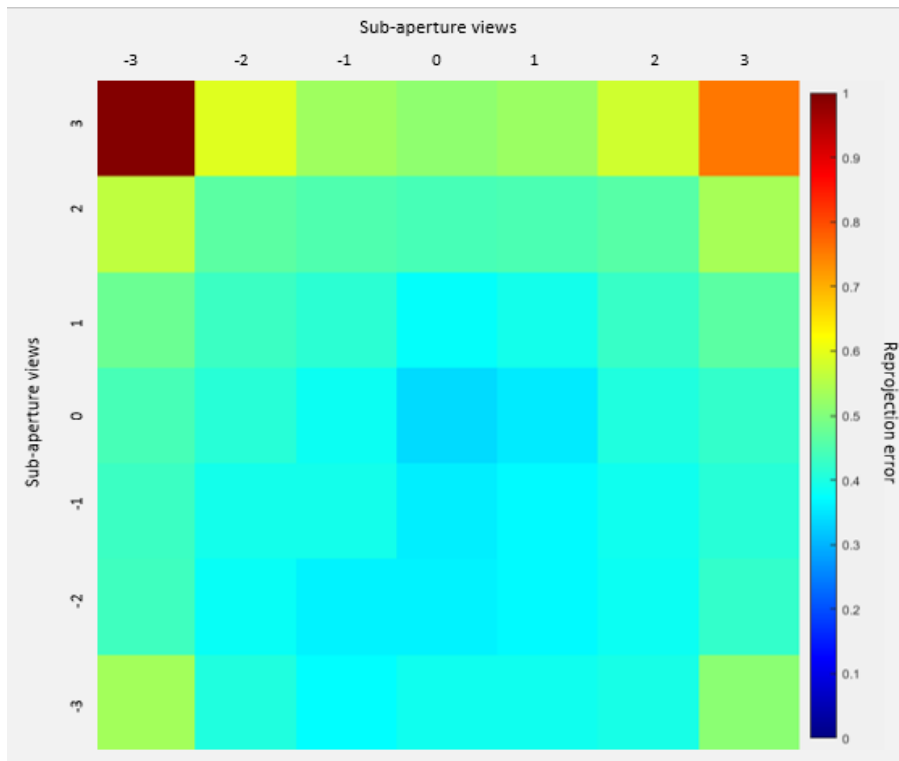


Figure 7.18 Mean re-projection error of sub-apertures using 35.1 mm grid generated using previous methods

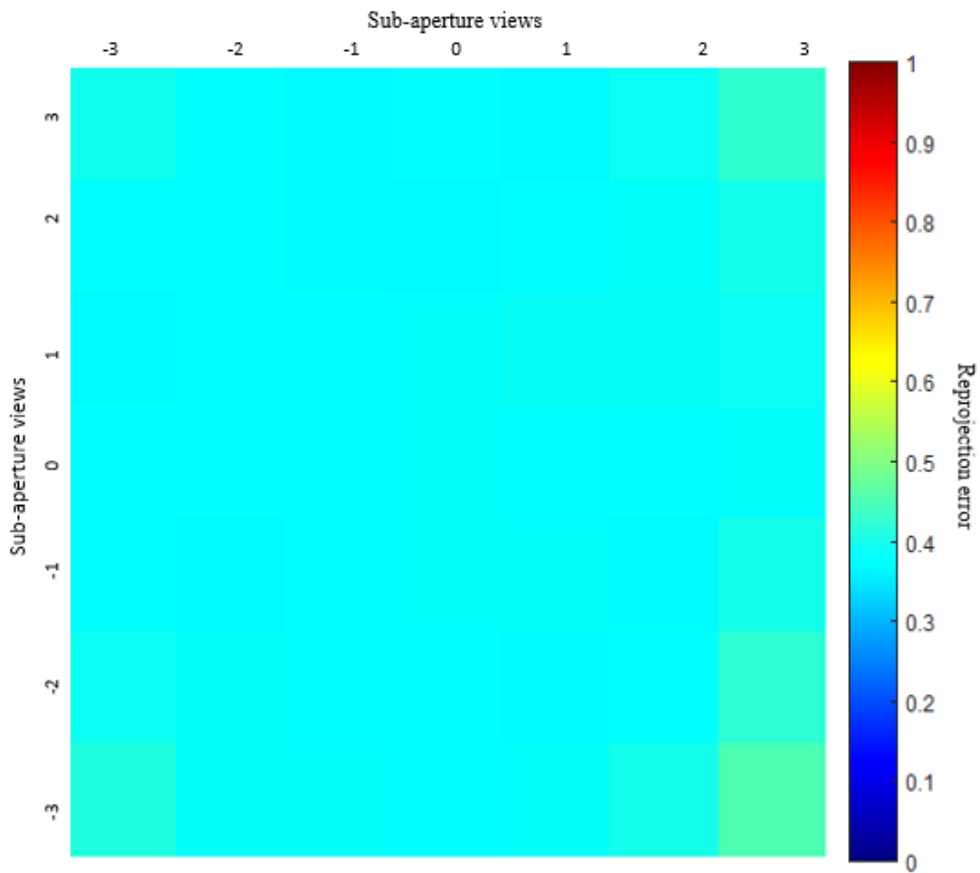


Figure 7.19 Mean re-projection error of sub-apertures using 35.1 mm grid-Lytro-I

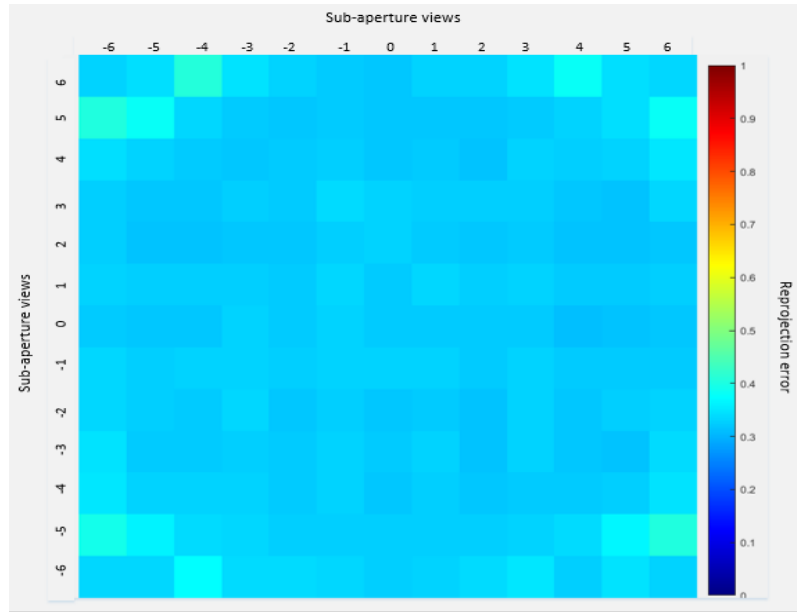


Figure 7.20 Mean re-projection error of sub-apertures using 35.1 mm grid-Illum

The overall calibration results are represented by calculating the disparity estimation between central view and the corner views of the Lytro camera data. Figure 7.21 shows the disparity estimation the Lytro cameras sub-aperture views before calibration. Figure 7.21 (left) shows the top view of the subject (toy) captured with a close range of 100 mm, while Figure 7.21 (right) shows the side view of the subject. The toy body and other features are at the same distance from the ground, whereas the gun and face are at different distance from the surface which has been distorted in the disparity data.

Figure 7.22 shows the calibrated data of the Lytro cameras using the technique illustrated in this chapter. It can be observed from the top view (left, Figure 7.22) that it has a close resemblance to Figure 7.21 (left) top view. While Figure 7.22 (right) exhibits the corrections and the effective results in the side view of the toy. The face and the gun are at different distances from the surface, and the 3D point cloud appears with less noise as seen in Figure 7.21 (right). Since the disparity values cannot be indexed using MeshLab software, thus disparity values are shown in Figure 7.23 as MatLab 2D plots. The results of the LDS and the calibration algorithm illustrated in this chapter can be observed by comparing the two sub-images in Figure 7.23 representing the raw sensor data (left) and raw distortion-free image (right) respectively. The distorted image obtained from the Lytro formatted file using LFtoolbox has been processed and the results are represented in Figure 7.23 that exhibit successful removal of distortion.

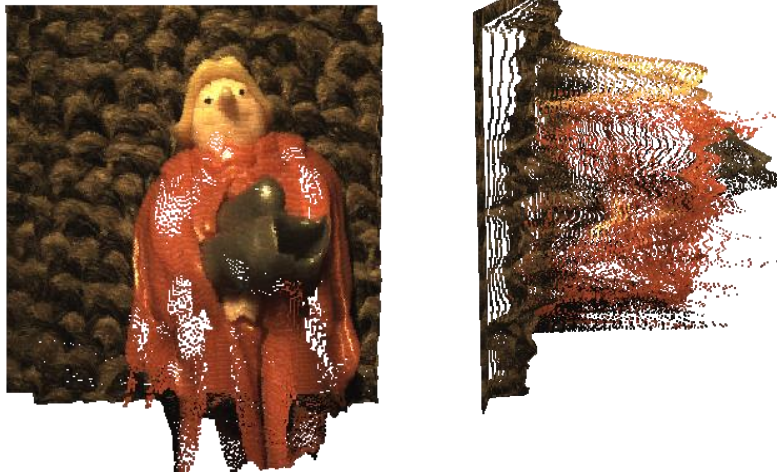


Figure 7.21 Disparity estimation of Lytro data before calibration, illustrated using 3D mesh using MeshLab software



Figure 7.22 Disparity estimation of Lytro data after calibration, illustrated using 3D mesh using MeshLab software

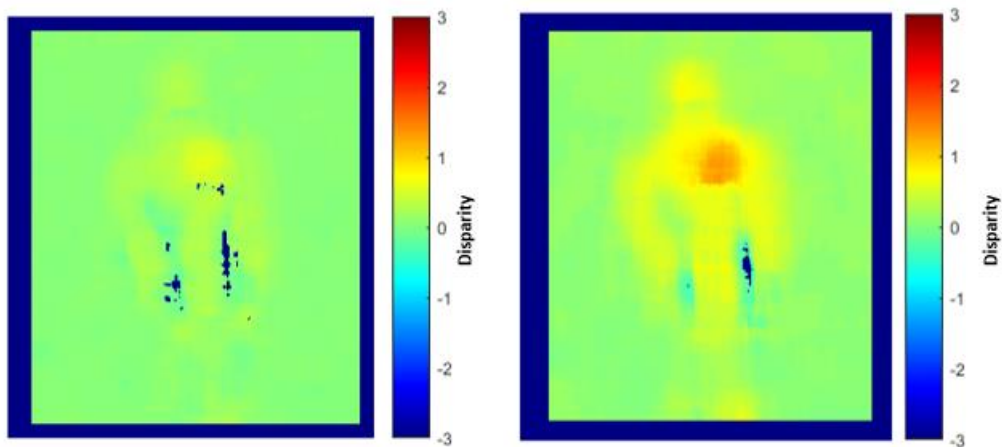


Figure 7.23 Disparity map of uncalibrated stereo images (left) and calibrated stereo images (right) used in Figures 7.22 and 7.23 respectively

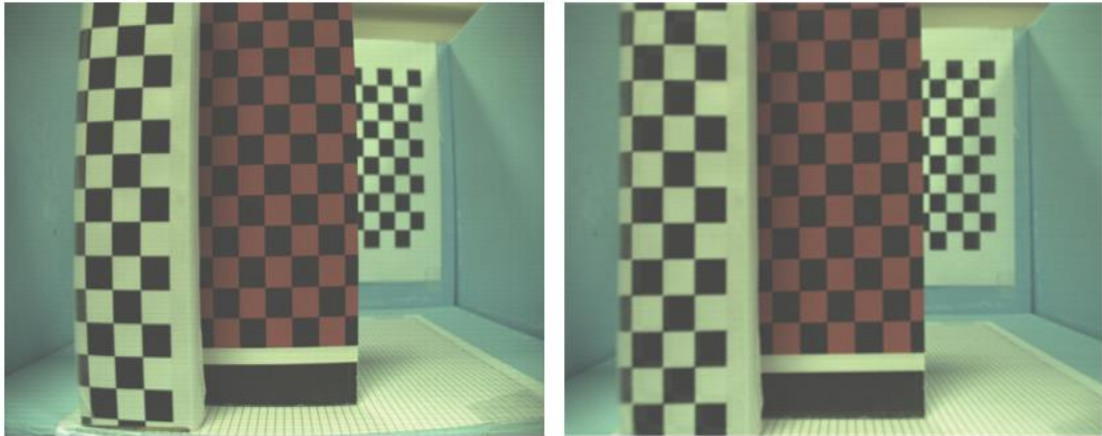


Figure 7.23 LF data after removing distortions

## 7.7. Conclusion

In the previous Chapters (5 and 6), the main source of LF data was generated by the LDS and the absolute depth map was calculated using the results given by the LDS. Hence the user has no control over the input source used for any image processing applications. This current chapter plays a key role to gain control over the LF depth generation process and provide flexibility overall parameters involved in generating input source.

This chapter illustrates initial steps to overcome the Black-Box effect, exhibited by the Lytro family of cameras. The Lytro cameras are the main source of generating LF data in this thesis, hence section 7.3.1 illustrates the method to acquire raw LF sensor images from the Lytro formatted files. Some of the initial observations necessary to determine the nature of the input sources and related corrections are suggested with necessary references. Since final results must be independent of error and distortions of the input source, a calibration procedure becomes necessary. This chapter illustrates a 2-step calibration algorithm that is based on the physical parameters of the MLA. Also, some of the key preferences necessary for calibration images and the method of generating sub-apertures images are illustrated with corresponding figures in Section 7.3 to 7.5.

A novel technique has been developed that utilizes the key parameters of the MLA to calibrate the LF cameras by using a virtual lens concept. This eases the process of calculating internal and external camera parameters along with quantifying the calibration result. In this technique, the key parameters such as; MLA pitch, microimage resolution, lightfield resolution and MLA cords, are used for the calibration

process. The new concept of the virtual lens reduces the lens complexity to a two-lens system thereby speeding up the calibration process. The MLA assembly provides a constant value between the sub-aperture images, and this value is used as a key to quantify the calibration results and provide values to update the calibration result.

Experimental results of calibration show lower re-projection errors on the entire sub-aperture images, which is approximately  $\pm 0.15$  pixel of the mean re-projection value (0.35 pixels) for the Lytro family of cameras. This method is efficient to generate calibration parameters in 64 minutes and 34 minutes for the Illum and the Lytro-I cameras with an image set consisting of 20 images while applying corrections in less than 1.2 minutes using a dual-core processor. This is approximately 50% faster compared to the competing algorithms, encoded and run on the same computer.

Furthermore, the calibration algorithm creates a stable platform to generate depth maps for LF data gathered using any MLA based LF camera. Here, the Lytro play the role of LF data supplier but has no effect on the depth results which was difficult in earlier chapters. Based on the results obtained from this chapter, the absolute depth maps are generated in Chapter 8.





# DEPTH ALGORITHM



## **Overview**

In Chapters 5 and 6, the images used for depth calculation were generated by the LDS. Since the user has no control over the LDS parameters or the technique used to generate results from the LDS is known, mapping the ‘input-process-output’ cycle becomes difficult. Hence to overcome the Black-Box Effect, a new platform has been designed that depends on the Lytro cameras for LF data and consequently the results of the process run independently of the Lytro cameras.

In this chapter, the methods suitable for depth calculation are related to the features present in MLA based LF cameras that support depth measurement are summarised. The main reason why the Lytro cameras have previously not been considered for engineering applications and the lack of absolute depth measuring technique is explained in this chapter. All the information collected for the MLA in Chapters 3 and 4 are used for the calibration procedure as explained in chapter 7 and the resulting software engine is represented in this chapter.

## **8 Absolute Depth from Lytro Camera Light Field Data**

### **8.1. Introduction**

The main aim of this work is to provide an absolute depth generating model that exploits the LF features available in the low-cost Lytro cameras. It is important to remember that, these cameras are manufactured with the intention of targeting the photography community. So, the camera design will be determined to meet the specification and features required by consumers of the photography community.

To set the depth algorithm free from the LF data capturing device, it is essential to understand the nature of the cameras and build a software package that provides results as required by the user. Chapter 4 provided the essential details with respect to the MLA parameters and methods to calculate the necessary details. While Chapter 6, illustrated the performance of cameras and introduced the TTF and CLS concept. Chapter 7 represented the key step of calibration, that enabled the users to generate distortion-free raw images from the Lytro camera. Using this information, a combined algorithm represents a package that generates results independent of the Lytro camera.

Before using the Lytro cameras for depth estimation, it is necessary to understand the key features of an LF camera that supports the results to be generated in absolute units. These features (for depth measurement) allows the user to select an appropriate method depending upon the applications.

### **8.2. Selection of depth calculation method**

The LF camera is very similar to conventional cameras with respect to design and the optical elements used to capture the data. To capture 3D data using conventional cameras, the location of the camera or the photosensor must be shifted by a known distance to calculate the depth information. Also, by using two or more conventional cameras and recording the disparity results in 3D data of the object space.

The LF cameras exhibit additional depth measuring features, beyond that of the conventional camera techniques. This is possible because the data recorded by the LF cameras can be transformed into individual camera views, as explained in Chapter 7. Depending upon the available camera data and the application requirements, the user can switch between three depth calculation methods such as; Ray Projection method, Disparity Estimation method and Focus Variation method.

### 8.2.1. Ray Projection method

The ray projection method considers the optical property exhibited by the MLA. In this method, the light rays are retraced to measure the 3D depth data recorded by the LF cameras. Figure 8.1 illustrates the ray projection technique by considering two light rays,  $a$  and  $o'$ , that are recorded on the photosensor pixel location indicated by  $x$  and  $y$  respectively.

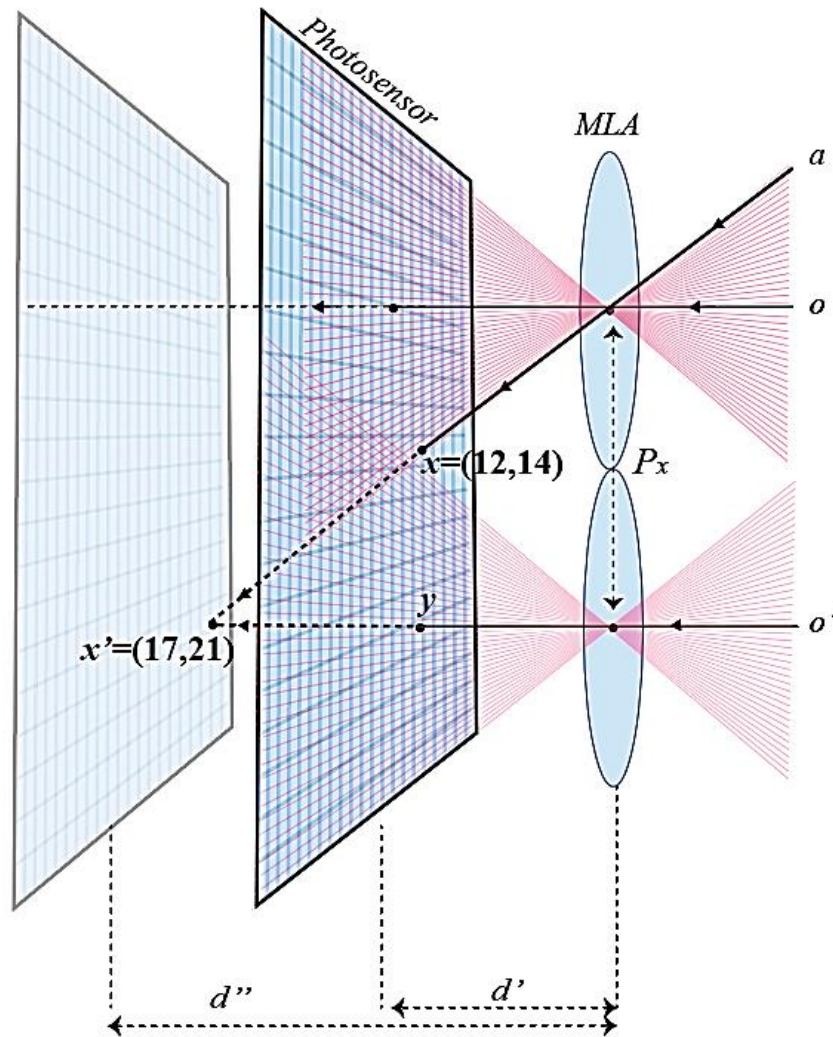


Figure 8.1 Refocusing technique and ray tracing using two lenses

This method assumes all camera design data are known, such as; MLA-photosensor assembly gap ( $d'$ ), the focal length of the camera ( $f$ ) and MLA cord details. By using feature detection techniques (e.g. RANSAC, SURF) [125, 126] the corresponding features between two lenses are detected and the ray tracing technique is used to calculate the depth.

The distance from the photosensor where the two similar feature points meet can be calculated by Equation 8.1 and the result is then used in Equation 8.2 to identify the true distance in absolute units. In Equation 8.1,  $p_x$  is the MLA pitch distance or the distance between the parent lenslet of the light rays considered and  $(x - y)$  represents the pixel difference of the common features under consideration.

$$d'' = \frac{p_x d'}{p_x - (x - y)} \quad 8.1$$

$$\frac{1}{f} = \frac{1}{z_0} + \frac{1}{z_1} \quad 8.2$$

This technique is suitable for the LF cameras where all camera data are available. When considering the Lytro cameras, many design features of both versions of Lytro cameras are not available, especially the MLA-photosensor gap ( $MLA - Ps, d'$ ). Hence the ray projection method was not considered as a potential depth measuring feature of the Lytro family of cameras. If  $MLA - Ps$  detail was available, the cameras could have been used to measure 3D data as exhibited by LF cameras designed for engineering applications [134].

### 8.2.2. Disparity Estimation method

One of the common techniques used in the engineering community to capture depth is by using multiple conventional cameras and measuring disparity between two camera views. Using conventional cameras, this technique requires alignment of camera views, focal length and baseline between multiple cameras. To overcome the effort of precise alignments, many software alignment methods are used by the engineering community by exploiting the principles of epipolar lines, feature detection and many other techniques.

Since the image captured using MLA techniques can be transformed into individual cameras views, a similar technique used in conventional cameras can be used for the Lytro cameras. The details necessary for successful estimation of 3D object space represented by Equation 8.3, where  $b$ =baseline between cameras, the  $d$ =disparity between two features (pixels) and  $f$ =focal length of the camera or resulting focal length of CLS.

$$z = \frac{fb}{d}$$

8.3

Further, the details necessary for depth estimation can be retrieved from the Lytro camera using the metadata file and by conducting experiments, and hence this technique is suitable to be used for depth calculations using low-cost LF cameras.

### 8.2.3. Focus Variation/Post-Focusing

Focus Variation is another important technique used in the engineering community, especially in metrology, where the camera is moved away from the object of interest at a regular interval and a stack of images are captured. By maintaining consistency between intervals and recording every interval in terms of metric values ( $-z$  axes) helps to map a relationship between the blur generated by captured images and the recorded metric value. This relationship plays an important role in deciding the absolute values for the object of interest [135].

The feature enabling the focus variation technique in LF cameras is explained in section 8.3. The gap between the MLA and the photosensor ( $d'$ ) is again the important detail necessary for successful generation of the image stack using the data captured by the LF cameras. Due to the lack of  $MLA - Ps$ , focus variation technique is not directly considered for absolute depth generation. But the image stack can be generated based on the ratio between the absolute and virtual distances, which gives sufficient information to generate the image stack and hence indirectly support the data generated by disparity estimation. Hence post focusing technique is considered for to generate relative data of the object space.

### 8.3. Post-Focusing / Refocusing

Conventional cameras have been used in engineering applications for many decades and still pose problems to the user in terms of out of focus images. To capture a sharp image of the scene/ object of interest there are two main criteria to be fulfilled. Firstly, the focal length of the lens and the distance of the photosensor should match. Second, the object distance to the plane of the camera should meet the requirement set by the thin lens equation [131], given by Equation 8.2, where  $f$  defines the focal length of the cameras,  $z_0$  represents the distance between the camera plane and the focus plane, while

$z_1$  is the distance between the main lens and the photosensor. If the cameras consist of multiple lenses, then  $f$  represents the combined focal length of the lenses.

With the help of Equation 8.2, the distance of the object ( $z_0$ ) from the camera plane is calculated in absolute units (mm) with known values of  $f$  and  $z_1$  or vice versa, to calculate other parameters. In conventional cameras, to measure different  $z_0$  of a scene, the focal length must be varied and there are many practical applications of changing the focal length of cameras to map the object space in 3D [136]. In contrast, varying the distance  $z_1$  also results in mapping object space, i.e. altering the location of the photosensor. In both the scenarios multiple images and additional information of a change in camera parameters are recorded that results in the absolute measurement of the object space. This concept is illustrated in Figure 8.2, where the object distance and the photosensor distance match the requirement defined in the Equation 8.2. A second object can be mapped on the photosensor, when the distance is equal to  $(z_1 - \Delta)$  from the main lens, when the object distance is varied.

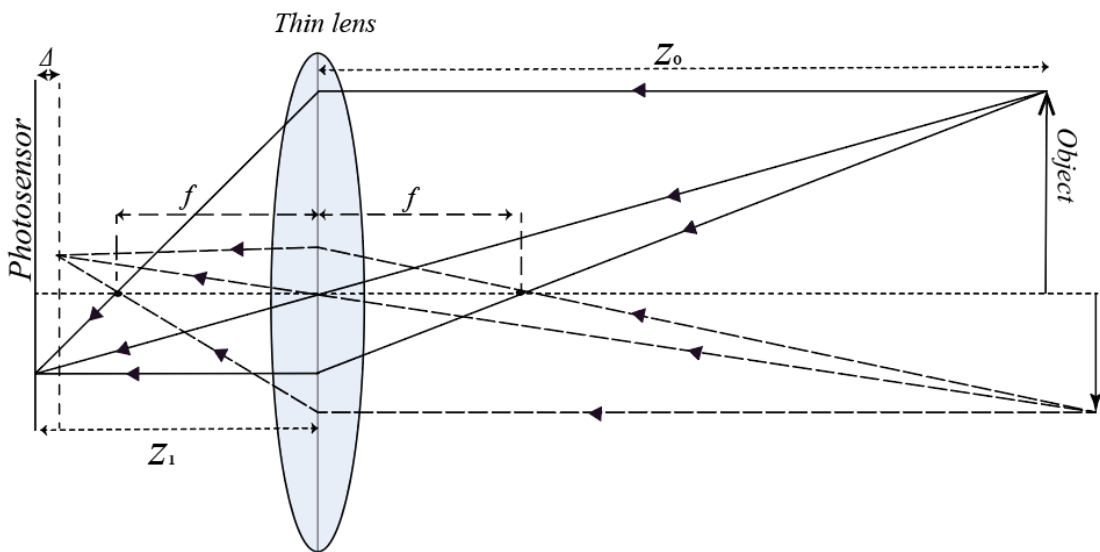


Figure 8.2 Necessary changes in a conventional imaging system to record different planes in the object space

As illustrated in Chapters 2 and 3, the LF cameras comprise additional optical elements that enable to record the direction component of light rays along with the intensity information. The refocusing formulae have been demonstrated with ray diagrams in Chapter 3 and the same concept is shown with actual values of the photosensors pixels and the MLA parameters. This will help to better understand the concept of refocusing using known cameras parameters.

$$x' = \left( u + (x - u) \frac{d''}{d'} \right) \quad 8.4$$

Equation 8.4 was illustrated in Chapter 3, so here the variables are replaced with practical values for demonstration. Figure 8.1 illustrates the scenario, where the photosensor is placed at  $d'$  distance from the MLA plane. Three light rays are considered,  $a$ ,  $o$  and  $o'$ , that pass through the MLA centre (known as the MLA cord, calculated in Chapter 4) and register on the real photosensor at a distance of  $d'$ . It can be observed that light ray  $o$  and  $o'$  align with the MLA centre and continue to travel along the same path. But the light ray  $a$  is recorded at location  $x = (12,14)$ , the pixel location with respect to the photosensor coordinate system. Until this point, all dimensions and values are real and hence represented with thick lines. To refocus the image plane, the light ray ( $a$ ) positions must be determined with respect to new photosensor location  $d'$ . Using known values, for example,  $d' = 25 \mu\text{m}$ ,  $d'' = 41 \mu\text{m}$ ,  $u = 3^{\text{rd}}$  pixel and  $x = 12^{\text{th}}$  pixel results in the  $x'$  pixel location as  $x' = 17.4^{\text{th}}$  pixel. The decimal values representing the pixel location indicate the necessity of interpolation between neighbouring pixel intensity or rounded off to  $17^{\text{th}}$  pixels from the MLA centre location  $u$ . It can be noticed that a second light ray  $o''$  also has the new pixel value of 17 and hence the integrated/combined intensity value is represented at the pixel location 17. The same procedure is carried in the  $y$  axis to obtain the new refocused image at  $d''$  distance.

$$\alpha = \frac{d''}{d'} \quad 8.5$$

To refocus the image plane, the distance at which the photosensor is to be moved virtually ( $d''$ ) and the real photosensor to MLA distance ( $d'$ ) must be known in absolute units ( $\mu\text{m}$ ). But, in Equation 8.4, it can be observed that the ratio of  $d''$  and  $d'$  are independent variables, the ratio can be replaced with a variable  $\alpha$ , that denotes the variable distance  $d''$  with respect to constant  $d'$ , as shown in Equation 8.5. The value of  $\alpha$  indicates the new location of the photosensor and it also indicates the new object plane location with respect to the object plane ( $z_0$ ) used to capture the image. From Equation 8.2, it can be observed that if the new sensor location moves towards the lens, then the new image plane will be behind the initial object plane and vice versa.

This points towards the important idea that, if  $\alpha$  is known or using different  $\alpha$  value the object plane can be mapped before refocusing. This eliminates the necessity of recording parameters of camera movements as seen in the conventional cameras. It can also be noticed from Equation 8.5 that to refocus at the new photosensor plane, the absolute values of the photosensor locations are not necessary, i.e. the ratio  $\alpha$  that has no units, is sufficient to refocus the image at different distances from the camera (virtually). But for a systematic refocusing (at known depths) the metric value of  $d'$  becomes very necessary. Figure 8.3 represents the refocusing algorithm results applied on a LF image captured using the Illum, where three checkerboards are placed at different distances from the cameras plane and refocused without using the actual metric value of  $d'$ . The ratio of distances,  $\alpha$ , has been used to refocus through different planes ( $\alpha = 3,2,1,0.1$ ).

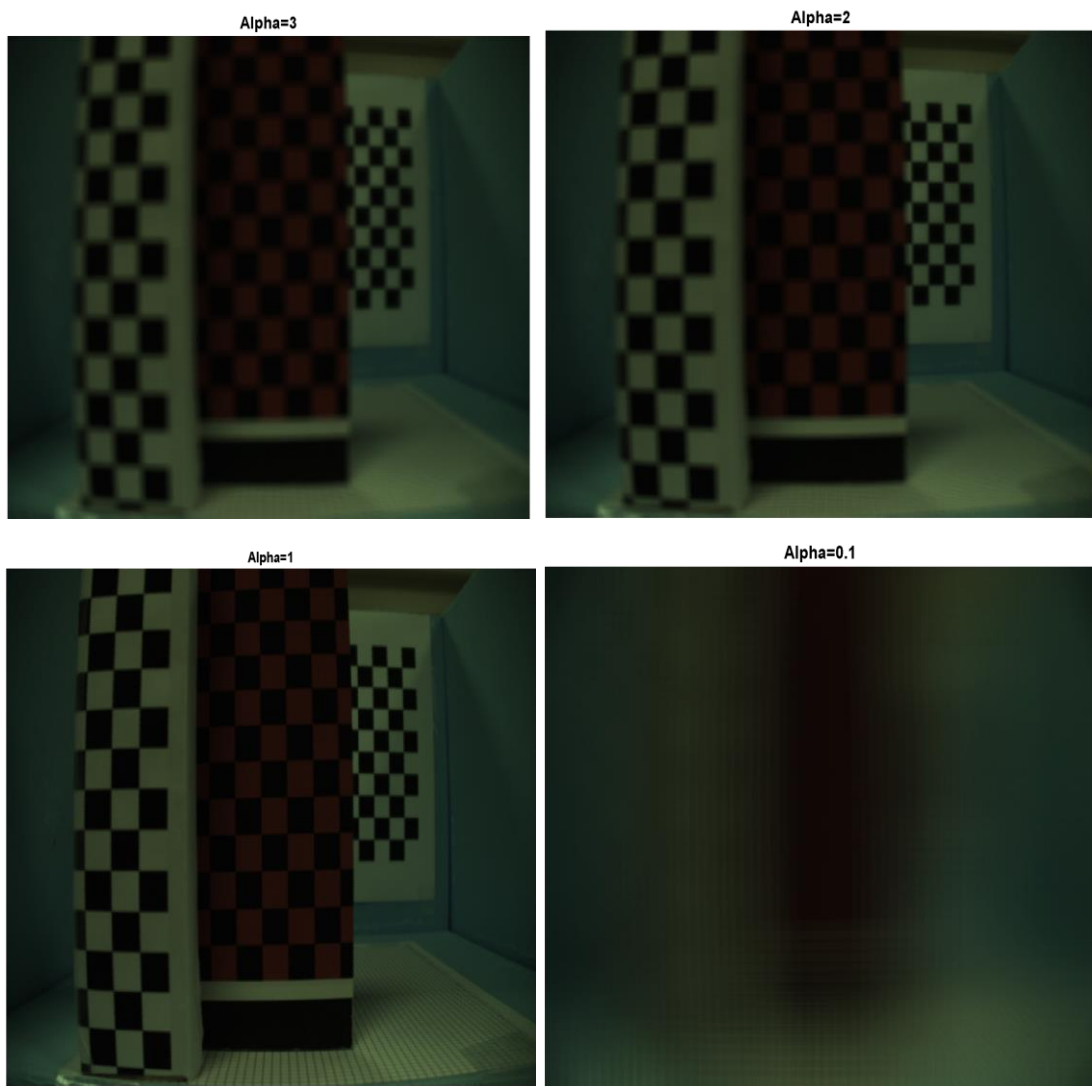


Figure 8.3 Refocusing at different plane/distances from the camera



The refocusing power of LF cameras has been previously investigated by using laboratory built test units, where the relationship between absolute refocusing (refocusing at a given object location) and camera design (complete details of camera components, lens parameters and distances) has been highlighted [27, 129, 130]. It was noticed that the Lytro-I cameras used in Chapter 5 had different build qualities and hence dismantling a Lytro camera and measuring the necessary details does not provide enough confidence to use the same parameters compared to other Lytro cameras for engineering applications.

Further, to refocus at different planes of the object space and generate image stacks the refocusing algorithm shown in Figure 8.4 can be used, where  $L_{mm}$  is the LF image in microlens addressing mode,  $L'_{mm}$  is the final LF image after processing. The necessary details required for successful relative stack generation are; MLA cord details, pitch and  $\alpha$ . The value  $\alpha$  is always represented as a product of  $\alpha_s$  which is the constant interval/space between two image stacks and hence  $\alpha$  is represented as  $\alpha = (\alpha_s n)$ , where  $n$  represents the total stack of images specified by the user. The method used to measure the alpha values corresponding to the focus/sharpness values is shown in the algorithm represented by Figure 8.5, where the gradient (line3) term defines the difference between adjacent columns and rows of the entire image.

---

**Algorithm: Refocusing**

---

1: Load LF image ( $L_{mm}$ )  
2: for  $x = 1: I_{cx}$   
3: for  $y = 1: I_{cy}$   
4:   for  $p_x = -p_x: p_x$   
5:   for  $p_y = -p_y: p_y$   
6:     $x' = (x + (x + p_x)(\alpha_s n))$     $y' = (y + (y + p_y)(\alpha_s n))$   
7: Interpolate any decimal pixel values  
8:  $L'_{mm}(x, y) = L_{mm}(x', y')$   
9: end all

---

Figure 8.4 Refocusing algorithm

---

**Algorithm: focus/sharpness**

---

- 1: Load LF image
- 2: for N=1: n
- 3: find gradient for  $\alpha=(\alpha_s \ n)$
- 4: apply an averaging filter

$$\frac{1}{p_x} \begin{bmatrix} 1 & 1 & \dots \\ 1 & 1 & \dots \\ \vdots & \vdots & \dots \end{bmatrix}$$

- 5: end
- 

Figure 8.5 Focus/sharpness algorithm

#### 8.4. Role of TTF in refocusing

In Chapters 6 and 7, the importance of the TTF has been illustrated. By activating the TTF feature, the user takes control over the focus plane and the camera's internal motors adjust the CLS (i.e. focal length) to match the user-desired focus plane in the object space.

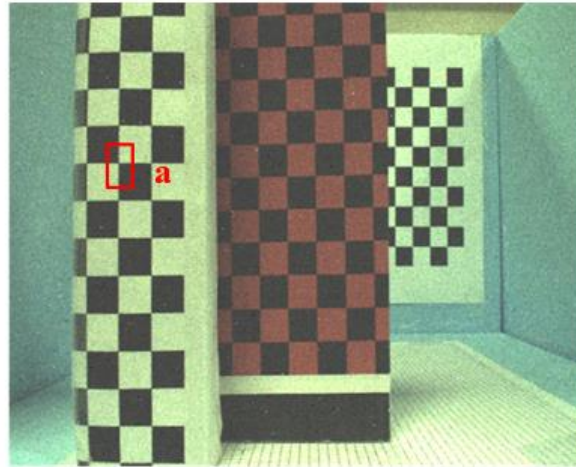


Figure 8.6 Image considered for alpha verses TTF experiment

The TTF feature has no effect on the performance of a given camera's ability to refocus at different distances. But the TTF provides additional information, helps to identify the focus plane with respect to the object space and divide the regions into different segments. The focus plane has an integral relationship with the TTF, as the focus plane results in high value at the TTF plane and declines on either side of the focus plane selected by the TTF (as defined in Figure 8.6 and 8.7). Figure 8.6 represents the image

considered for the alpha-TTF experiment with pixels within the square ( $a$ ) considered for sharpness measurement in all the stacks. In Figure 8.7, it can be noticed that the  $\alpha \simeq 1$  corresponds to the sharply focused plane as indicated by previous chapters, i.e. values of  $d'$  and  $d''$  differ slightly because the product of  $(\alpha_s n)$  do not match to 1 always. In Figure 8.7,  $\alpha_s = 0.0142$ ,  $n = 72$  out of 101 slices and the sharpness index was equal to 0.01259.

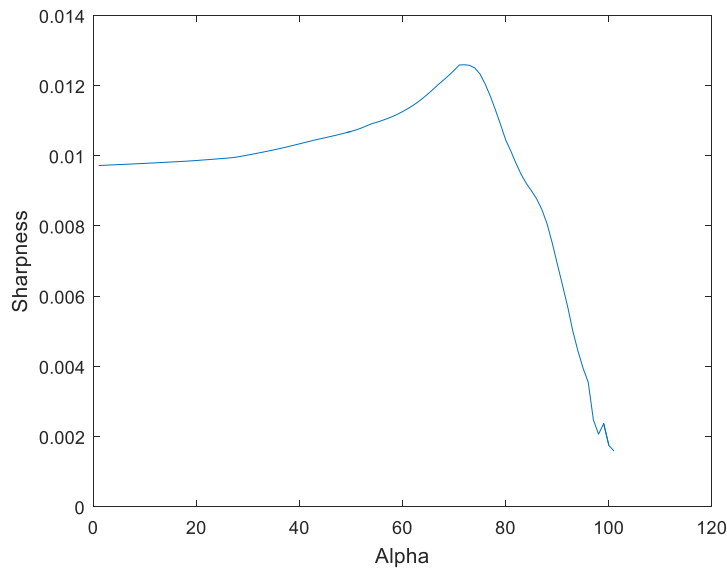


Figure 8.7 Alpha value compared with the TTF of the Lytro cameras

## 8.5. Depth algorithms

The depth measuring features available with the Lytro family of cameras are briefly explained in Section 8.3. Before the Lytro camera captures data and is processed for generating depth, the Lytro files must be decoded to obtain the raw sensor images. A detailed procedure starting from extracting raw sensor data up until the depth calculation is shown in Figure 8.8, where every box represents individual steps to be followed to generate absolute 3D data.

Some of the steps shown in Figure 8.8 have already been explained in the earlier chapters. A brief walkthrough of each step and the corresponding chapter details becomes important to understand the concept easily. The Lytro metadata file of the object space captured using the Lytro cameras are generated with the help of the LDS. The metadata file is then processed using the LFtoolbox and the Exiftool to generate raw sensor data and corresponding camera details (focal length), as explained in Chapter 6.

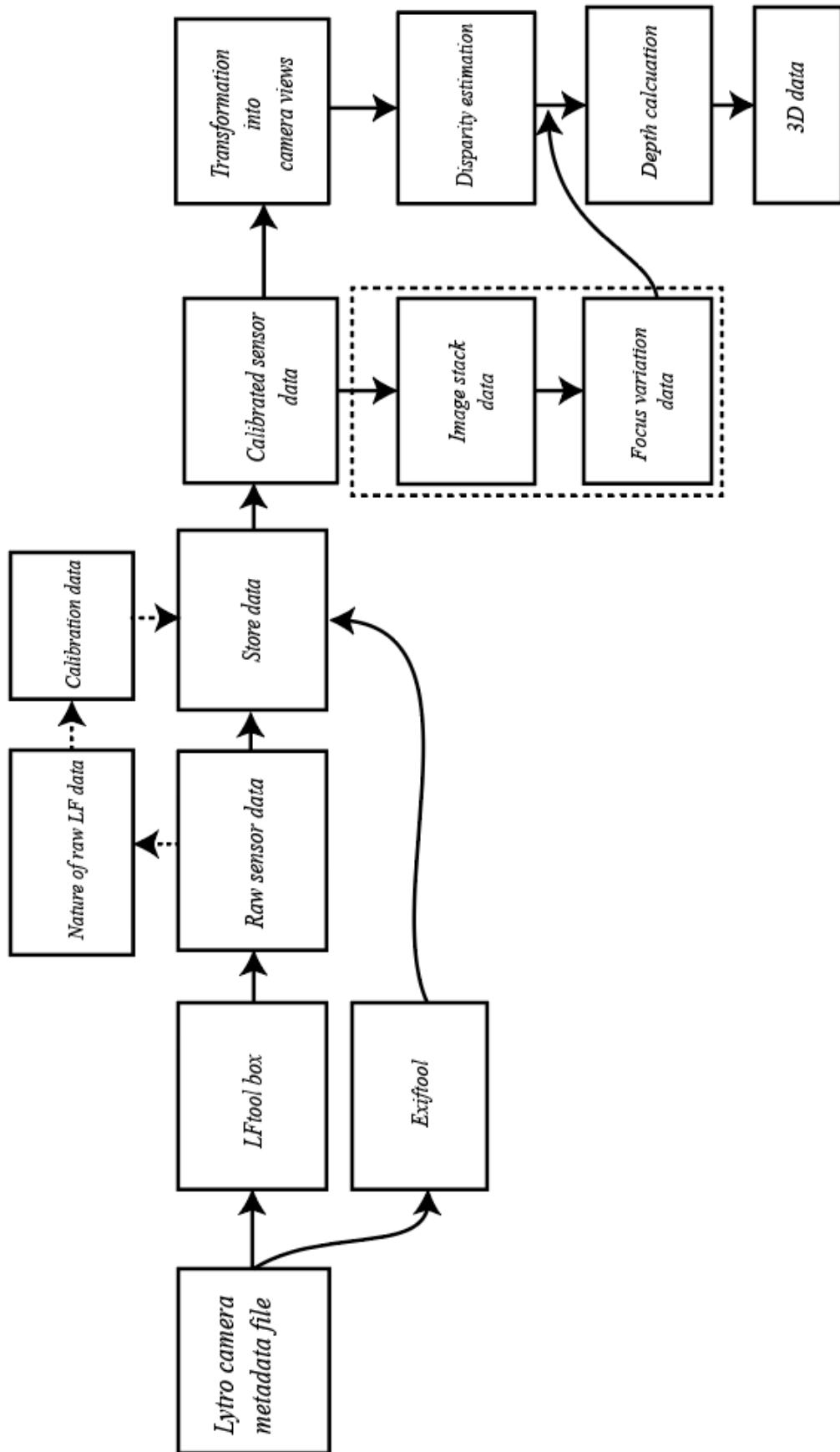


Figure 8.8 Procedure to generate absolute depth data using disparity method

The raw sensor images are later processed to remove any distortion using the calibration algorithm described in Chapter 7. Some of the pre-requisites, such as the MLA parameters and the MLA cord details (see Chapter 4), are calculated for the specific camera and are stored as required by the calibration algorithm and nature of the cameras are manually identified (see Chapter 6).

Further, the distortion-free LF data is processed to generate individual camera views that are required for disparity estimation (see Chapter 6). For generating additional data using the Focus Variation technique, the calibrated LF images are processed using a refocusing algorithm (Figure 8.4) and the image stacks are saved for blur estimation. Finally, using the disparity map in the Equation 8.3 the depth data is calculated in absolute units. If necessary, the additional blur /focus variation details of the object space can be matched with disparity data using graph theory (MRF) although this step is optional.

---

**Algorithm: Absolute depth using disparity method**

---

**1:** Load base disparity ( $b_d$ ), focal length ( $f$ ), user defined plane distance ( $Z_{mm}$ ) (as defined in chapter 7)

**2:** Calculate the disparity map ( $d_{map}$ ) using *PR image* as a reference ( $d_{map}/NH$ )

**3:** Vectorise  $d_{map}$  into  $d_{val(1...i)}$  with  $i$  elements

**4:** Calculate the absolute distance ( $d_{mm}$ ) for  $n$  elements

**for** ( $i = 1:n$ )

Case 1: ( $d_{val} > b_d$ ) & ( $d_{val} > 0$ )

$$d_{mm} = (b_d f / |d_{val}|)$$

Case 2: ( $d_{val} < 0$ )

$$val = (|b_d| f / |b_d - d_{val}|)$$

$$d_{mm} = 2 * val$$

**end**

**5:** Reshape  $d_{mm}$  to match  $d_{map}$  size to generate absolute depth map w.r.t reference plane ( $D_{mm_{ref}} = d_{mm}$ )

**6:**  $D_{norm} = (d_{mm}) / (f b_d)$

**7:**  $D_{mm} = D_{norm} + Z_{mm} - 1$

---

Figure 8.9 Algorithm for depth calculation using disparity method

Figure 8.9 represents the depth algorithm used to generate absolute depth using the Lytro captured LF data. The initial values required by the algorithm are specified in line-1 such as base disparity ( $b_d$ ), focal length ( $f$ ), user defined plane distance ( $Z_{mm}$ ). The method to obtain these values are illustrated in Chapters 6 and 7. The value  $f$  can be recovered using Exiftool, while the base disparity values are shown in Table 8-1 that was generated in the calibration step. The variable  $d_{val}$  is the transformation of  $d_{map}$  into a vector, i.e. conversion of row and column of image into single column.

Table 8-1 Base disparity corresponding to NH value

Camera	NH1	NH2	NH3	NH4	NH5	NH6
Lytro-1	0.26	0.61	0.78	0.0	0.0	0.0
Illum	0.48	0.94	1.32	1.80	2.12	2.56

The disparity map generated by considering  $PR$  images (see Chapter 7) as a reference, are divided by corresponding neighbourhood value to normalise the  $d_{map}$ . The image under consideration and the normalized  $d_{map}$  are shown in Figure 8.10, where the positive disparity represents objects close to the camera, while negative disparity represents the far distance. The disparity values are then converted to relative values (mm) with respect to the focus plane using the equations specified in line-4 of the algorithm and the corresponding relative plane depth map is represented in Figure 8.11. The relative depth map is generated by considering the focus plane as a reference since the disparity in that plane is equal to the base disparity of the corresponding cameras. Significant noise can be seen where a perfect pixel disparity match is not available. Figure 8.12 represents the top view of the result generated in Figure 8.11 illustrating the different boundaries of the depth map. By using the user provided focus plane distance, the depth map can be re-calculated to provide the distance of the object space with respect to the camera plane, as shown in Figure 8.13. Since the object space was measured in the camera independent version, the refocusing was achieved using manual values that range from  $\alpha=0.2$  to 2 with 256 image stacks. The Focus Variation for close

range measurements are not efficient [138] and do not add much detail to the disparity data and hence are neglected (see Figure 8.14).

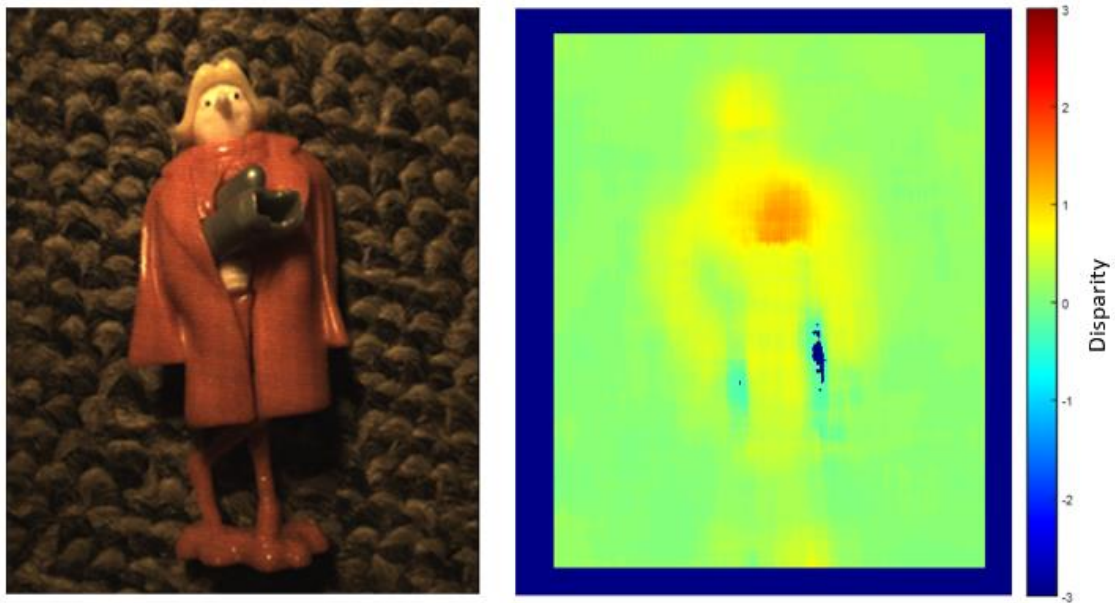


Figure 8.10 Normalized disparity estimation(right) and image considered (left)

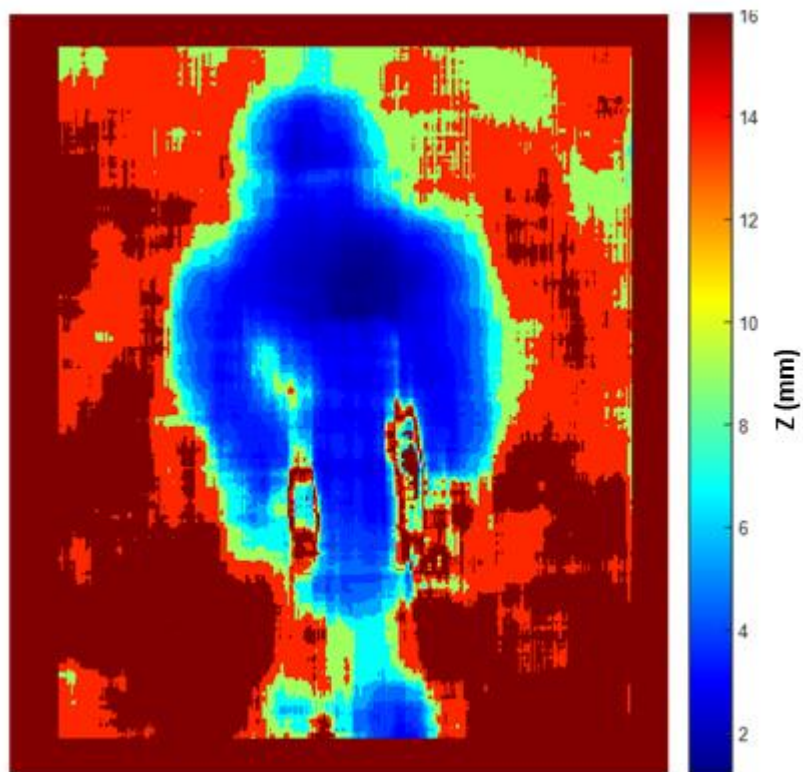


Figure 8.11 Relative plane depth map (mm)

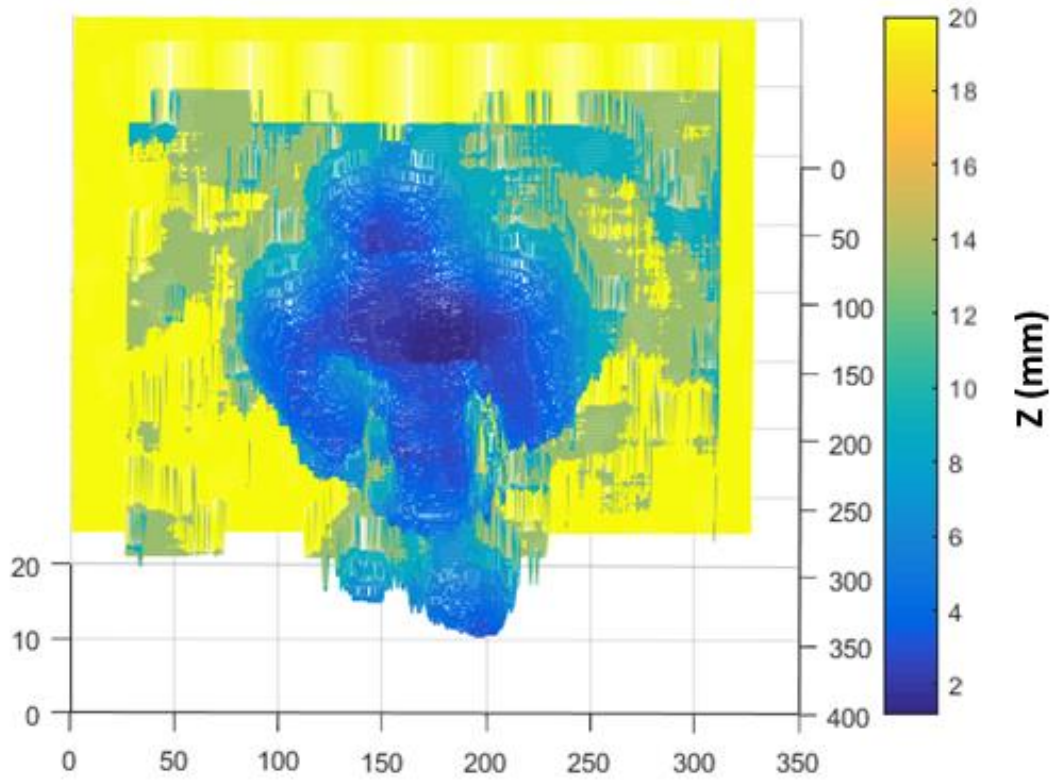


Figure 8.12 Top view of the result (mm)

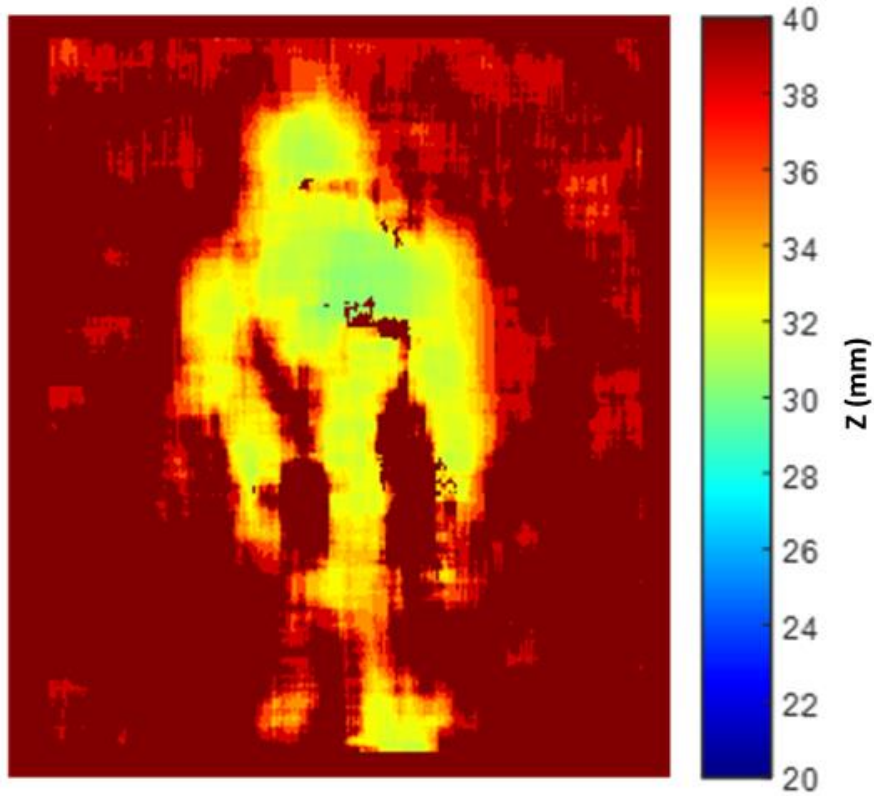


Figure 8.13 Absolute depth map relative to user-defined focus plane distance



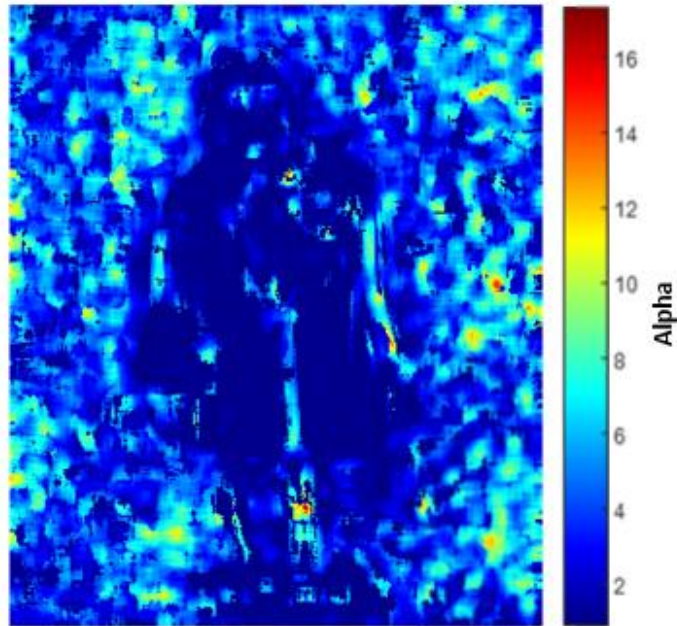


Figure 8.14 Focus variation results for different values of alpha

## 8.6. Conclusion

The data collected in Chapters 4 to 7 provide essential details about the Lytro family of cameras, that provides a platform to develop a MATLAB engine that works independently to the parent camera responsible for capturing the LF image. Since the Lytro cameras are the target devices in this thesis, the results presented are the raw details without adopting any further region merging algorithms.

It has been noticed that the MLA-Ps assembly details play an important role in generating and accessing features of the LF cameras. The MLA parameters calculated in Chapter 4 provide essential details that enable access the static details from the LF cameras, i.e. the details that do not require ray tracing. While other results such as refocusing, and depth calculation using ray projection methods require the fixed distance of the MLA-Ps assembly. Due to the lack of this information, the method adopted to achieve depth results are shifted to the disparity method where the MLA-PS assembly gap is not necessary.

The major distinctions between cameras that are specifically designed for engineering applications and commercially available cameras (Lytro) is the availability of specific camera details necessary for accessing LF camera feature. The Lytro cameras are built for the photography community provide details required by the users such as f-number, intensity values, shutter speed and ISO values. In contrast, details required for

engineering applications are not made available which prevents the normal use of the Lytro cameras practical for applications to measure the depth in absolute units.

It must be noted that the build quality of the Lytro cameras is not consistent, as seen with the Lytro-I cameras (LC1 and LC2). Hence any measurements specified for a particular Lytro camera are not necessarily representative of an entire group of cameras, which further suppresses the accessibility of Lytro cameras features. However, by using the details generated by the LDS results can be provided that are helpful to generate depth values in absolute units (Chapter 5 and 6).

Hence with all these limitations, a camera independent MATLAB engine helps to access the disparity details within a short range of 100 mm from the cameras. The results provided in this chapter illustrate the depth measuring accuracy in the order of a few millimetres with sufficient features to detect the matching pixels. The algorithm and the base disparity values generated in the calibration steps help for calculating the depth values in absolute units exhibiting the novelty of this work. Some of the major drawbacks of the Lytro cameras and corresponding feature to establish depth detection have also been highlighted in this chapter.

The overall observation and results witnessed during this work in correspondence to the commercially available Lytro cameras are presented in next chapter, where important results and their novelty are brought together as conclusion statements.



# 9



## CONCLUSION & FUTURE WORK

## **9 Conclusion**

### **9.1. Introduction**

The research work presented in this thesis represents the results and goals achieved with respect to the objectives and aims identified in Chapter 1. The two core objectives that acted as the driving force of this research and thesis were:

- To understand the concept of LF imaging from an engineering and hence a metrological point of view, along with acquiring knowledge regarding the principles of LF imaging and essential components used (MLA).
- Use a commercially available low-cost, microlens array based LF camera (Lytro) to familiarise and obtain hands-on experiences with LF cameras and identify the potential suitability of these cameras for engineering applications.

Both of these aims, along with the additional results that were identified through experimentation are summarised in Section 9.2 and Table 9-1 highlights the key contributions and novel elements categorised according to the chapter list, along with future work highlighted in Section 9.3.

### **9.2. Key Conclusion**

With reference to the aims and objectives identified in Chapter 1, key conclusions based on this research are presented as follows:

- This work has demonstrated a detailed understanding of the LF imaging technique. The LF technique has been presented from a bottom-up approach to ensure that the key accepts of this imaging technique are understood. The ray diagrams and light ray models illustrated in Chapters 2 and 3 uplifts the level of understanding required for analysing the LF cameras.
- This study has compared LF and conventional camera techniques in most of the chapters by highlighting the key differences and similarities between them. The main features of LF imaging that overcomes the limitations of

other conventional imaging techniques along with the component responsible for it are explained.

- A key distinction of this work has been with the understanding and representation of the interaction of light rays with the MLA. As mentioned in Chapter 3, mathematical equations may be difficult to understand for someone new to the LF area without the context of appropriate optical theory and modelling. Hence a major portion of the work has described the interaction of light with the MLA by considering fewer light rays. This has been supported by ray diagrams corresponding to relevant theory.
- Also, the main concept of LF imaging is based on the pinhole camera model. This has been exhibited in Chapter 3, where two light ray models have been considered for illustrating the key advantages of LF cameras that follow the primary ray model, while conventional cameras follow a secondary ray model and hence fail to record the necessary depth details for post-processing.
- A major portion of this work concentrates on the interactions of camera elements to produce LF images. The hardware requirements for the LF cameras have been specified in early research, yet the specifications of pixel size, focal length and resulting outcome have been investigated to highlight the hardware requirement necessary depending upon the applications in Chapter 3.
- One of the aims of this work was to target low-cost LF cameras models and determine their suitability for engineering applications. Hence, Lytro cameras were considered for experimentation and verification because they are low-cost cameras (typically £ 100 to £ 1,000), and work based on the MLA

technique. Hence a detailed analysis of the Lytro family of cameras has been carried out using two version of the cameras, the Lytro-I and the Illum cameras.

- The most basic details necessary for using any MLA based cameras are the MLA cord information. This was defined in Chapter 4, where a general approach of designing algorithms can be seen, i.e. the algorithms are designed for the Lytro family of cameras, but the algorithms are also designed to suit any MLA based LF cameras. In addition, these algorithms are developed and employed to perform automatically to detect the MLA cord without depending on user input. This has been a key development where no details of the MLA are required to generate the MLA cord information.
- A new distinction has been identified where previous research results use the geometrical system to find the MLA cord, while in the practical scenario the light rays behave differently as explained in Chapter 4. A light directed method has been developed where light rays follow the pinhole model.
- The primary evidence suggesting the suitability of the Lytro camera in engineering applications is presented in Chapter 5, where the camera generated results are used to map the relationship between absolute distance and the greyscale values. Also, new characteristic features of the Lytro family of cameras were identified such as; light sensitivity, depth response to the colour of the object, and pixel resolution in terms of absolute scale. With these results, the working zones for the Lytro family of cameras were defined such as the Active Zone (*AZ*) and Inactive Zone (*IAZ*). The work zone results suggest that the Illum cameras exhibit higher *AZ*, approximately 650 mm, while the Lytro-I cameras (both LC1 and LC3) exhibit *AZ* of 200 mm with

both cameras showing depth accuracy of  $\pm 10$  mm. The information provided in Chapter 5, uses the repeatable nature of the Lytro results to calculate absolute depth.

- A comparison of two versions of Lytro cameras, the Lytro-I and the Illum, have been made for the first time in an engineering context. The depth and colour sensitivity of both the cameras were explored using different means such as; varying light intensity under laboratory condition and using a uniform light source to identify the response to the change in colour. The result suggests that the Lytro cameras greyscale data are suitable for objects with low varying colour contrast with one or many (max 4 objects tested in this work) objects in the field-of-view and are especially suitable if all objects are of the same colour. However, objects with high colour contrast can cause incorrect depth measurements results.
- The Lytro family of cameras were compared one-on-one for their performance in identifying user-selected object/ Z plane. The experimental results suggest that the Illum camera has a high accuracy of adjusting the internal lens system to select the user desired object. While the Lytro-I occasionally performs well when the object is close to the camera plane, with performance declining as the object moves further to the cameras. The high confidence with the Illum cameras is due to the complex lens system (the Lytro-I lacks this feature). The confidence with the Illum cameras is reliable under 1,000 mm of the object distance, measurement beyond 1,000 mm was not determined in this thesis. This being the first evidence-based results for the Lytro family of cameras in metrology context.



- In addition to the camera performance, the Lytro stereo data was mapped to the original hexagonal structure of the MLA. This information plays a crucial role to prepare the way for the use of the Lytro cameras in engineering applications. Because the mapping details are known, the disparity estimation can be compared with the standard disparity (base disparity) and hence generate the disparity results in absolute units.
- A flexible work volume of 100 mm on either side of the focus plane was identified for the Illum cameras while using stereo data. This suggests a total of 200 mm of work volume that moves along with the focus plane within the 1,000 mm of the experimental region. With the algorithm suggested and known base disparity, the Illum cameras can be used to measure depth in absolute scale and this method is not limited by a change in illumination or the shape of the object. An additional method is suggested to take advantage of the disparity generated by the Lytro-I cameras with user-provided base disparity data to estimate the depth details in absolute scale.
- To overcome the Black-Box Effect and consider the Lytro family of cameras as an input source of LF data (but has no effect on depth calculation), raw sensor images were extracted and calibrated. The MATLAB LFtoolbox was used to extract the raw information from the Lytro coded metafiles. A novel calibration method was developed to involve MLA details in the calibration process by following virtual lens techniques and double pinhole methods, resulting in 4D calibrated data.
- This calibration method produced results in close agreement with existing methods, with a better performance on all sub-aperture views generating an error of  $0.32 \pm 0.15$  pixels for the Lytro family of cameras. The novel

algorithm also provides a qualitative measure of calibration using a neighbourhood method. This also indicates the sub-aperture data that can be used before applying the complete calibration procedure depending on the applications.

- The calibration method developed here provides the results 50% faster compared to other existing methods when processed on the same PC.
- The available depth measuring technique (disparity method) was explored and a Lytro camera independent platform was designed where the algorithm depended on the Lytro cameras for the input image. The algorithm generated two sets of results. Firstly, the absolute depth map is relative to the user selected plane (using the TTF), i.e. the depth data varies with respect to the plane in object space. In this situation, depth variation is independent of the camera position. The second result depends on the user input data; hence this depth data can be relative to user detail and accuracy of user information.
- A successful, absolute depth detection algorithm was designed using the base disparity between the camera views. The algorithm was capable of generating absolute depth of small object of size measuring in the range of 50 mm x 15mm (height x width) with sufficient features to detect a disparity between cameras views.

The overall conclusion from this work suggests that the Lytro cameras being cheap exhibit features that are helpful in engineering applications. By using the results from the LDS, long-range measurements are possible ranging from 100 mm to 1,000 mm (see Chapter 5 and 6) with accuracy up to  $\pm 10$  mm. By using the Lytro independent LF processing method, due to the lack of important MLA details the working volume has to be approximately within 100 mm. The Lytro cameras are suitable for engineering applications where the accuracy required are in the order of millimetres and with

moderate complex shapes. The depth results depend on the quality of the image captured, and consequently, sufficient lighting is necessary for good depth results.

A brief overview of the novel work executed in the of 3 years research is tabulated in Table 9-1, with an estimation of the extent of novelty (0-5).

Table 9-1 Novelty rating corresponding to individual chapters

<b>Chapter number</b>	<b>Chapter and novelty description</b>	<b>Novelty rating (0 to 5)</b>
2	<b>Review of Light field cameras</b>	0
3	<b>Light Field Acquisition Using Microlens Array Based Cameras</b> <ul style="list-style-type: none"> <li>• Ray diagrams explaining light ray models in LF cameras and pinhole concept</li> </ul>	1
4	<b>Parameters of Microlens array</b> <ul style="list-style-type: none"> <li>• Automatic detection of the MLA cord without user inputs</li> <li>• Identification of light ray directed MLA cord system</li> </ul>	3
5	<b>Absolute depth using Lytro cameras depth map data</b> <ul style="list-style-type: none"> <li>• Identification of the relationship between the Lytro camera depth map and illumination</li> <li>• Developing database relating absolute depth (mm) to greyscale depth data of Lytro camera (8-bit) for 3 cameras under two lighting conditions.</li> <li>• Mapping spatial pixel resolution to absolute units (mm)</li> <li>• Identification and using work volume of the Lytro cameras and classification of field-of-view into Active and Inactive Zones</li> </ul>	5

6	<p><b>Absolute depth using stereo view data from Lytro cameras</b></p> <ul style="list-style-type: none"> <li>• Identification and mapping of Lytro perspective data corresponding to MLA structure of the Lytro cameras</li> <li>• Identification of touch-to-focus (TTF) feature in correspondence with the performance of Lytro cameras</li> <li>• Identification of base disparity between perspective views in the Lytro cameras and using it to quantify the quality of complex lens systems in the Lytro cameras.</li> <li>• Developing an algorithm to identify the absolute depth using base disparity data and disparity generated between perspective views</li> <li>• Measuring the flexible work volume that corresponds to the focus plane governed by the TTF feature.</li> <li>• Evidence-based results pointing towards the quality of data generated by the Lytro family of cameras</li> </ul>	5
7	<p><b>Calibration of microlens array-based cameras</b></p> <ul style="list-style-type: none"> <li>• Identification of dual views and reporting the same corresponding to the Lytro cameras.</li> <li>• Developing a novel calibration system using a double pinhole camera model.</li> <li>• Detailing the calibration image requirement for better results</li> <li>• Identification and developing the corresponding algorithm to detect base disparity and using the same features to calibrate the stereo lenses thereby enhancing the overall results of calibration.</li> <li>• Improvement of the reprojection error over different views of the Lytro cameras and speeding up the calibration process by 50% using the same amount of image sets, compared to other algorithms.</li> <li>• Identification of a qualitative approach to</li> </ul>	5

	measuring the calibration results using neighbourhood function.	
8	<p><b>Absolute depth using Lytro light field data</b></p> <ul style="list-style-type: none"> <li>• Reporting the main reasons for preventing the Lytro cameras to be used for engineering applications</li> <li>• Listing the possible ways to generate absolute depth using the Lytro cameras</li> <li>• Development of an algorithm to exploit disparity between sub-aperture images and calculate the depth in absolute scale.</li> </ul>	5

### 9.3. Future work

The overall aim of this thesis was to identify the potential features of LF cameras, especially the Lytro low-cost cameras to record and generate 3D coordinate data in absolute scale. The work has pointed towards the features of the Lytro cameras that can generate absolute depth. The future work section illustrates theoretical and applied research directions that would naturally extend from this reported research.

Few important directions pointed by this work are:

- The main characteristic features identified with the Lytro cameras such as; work volume, colour and illumination sensitivity affecting final depth results, should be applicable to the LF cameras designed for engineering applications. Hence further research similar to the Chapter 5 experiments needs to be carried with Raytrix or other engineering LF cameras to understand the overall nature of LF cameras.
- The research carried out here represents the Lytro cameras through the raw sensor data available through the LDS or the LFtoolbox, and additional information collected related to the MLA. For an extended analysis of the Lytro cameras, experiments with different parts of the Lytro cameras are suggested, i.e. conduct a series of experiments with the lens elements used in

the Lytro cameras to provide an enhanced understanding of these low-cost cameras.

- The zoom feature of the Lytro cameras has not been explored in this research, hence conducting experiments by considering the zoom feature may shed light to unexplored areas of the Lytro imaging in the context of engineering applications. This is important because if a camera function is available-users will use it, with little understanding of the consequences to the integrity of data.
- The research represented here considers the only static scene of the object space. A natural extension of this work would be to identify the effect of real-time (dynamic) LF imaging and generating useful information to suit different applications in the engineering domain.
- The algorithms provided in this research can be optimized for higher efficiency with respect to the time complexity. For production environment applications, converting the MATLAB code to other platforms (e.g. C++) would boost the usability of the Lytro cameras in engineering applications.
- Minimal results have been identified with respect to the shape of the MLA effecting/supporting LF data capture. In this research, the importance of the active surface of the MLA has been described using the hexagonal shape of MLA lenslets. Therefore, for future developments and to understand the importance of different MLA shapes and structures, experimentations needs to be carried with square, circular and other shapes become necessary.
- Similarly, the aperture size, photosensor pixel size and the lens elements have been considered in this work in a broad perspective. Hence, research to generate a relationship between the optical lens components with the hardware elements would provide more information to help to build application-specific cameras models.
- The LF cameras are not widespread in the field of engineering applications and one of the reasons is related to the work volume. The few existing results, along with the detailed results from this research suggest a limited work volume approximately 1,000 mm. Hence, experimentation based on the current results should be extended to large volume/large objects to identify the

suitability of LF cameras, especially the low-cost Lytro cameras, in large area measurements.

- The results from the current research presented here, provide sufficient evidence and method to suggest applicability to use the Lytro cameras for microscale/microscope related applications, i.e. using the Lytro cameras for applications for measuring flat, less complex shape object that fit under a microscope or close to the camera exit pupil.
- In this research, a basic object with low complexity in shape has been used such as; flat objects and toys with the less curved surface. Hence, an investigation with complex shapes would provide information to fill the knowledge gap in the field of LF cameras in engineering applications.
- Conducting experiments to generate response curves with higher resolution, i.e. in this work, the resolution of response curves is around 5 mm and hence an error of  $\pm 5\text{mm}$  is exhibited by the depth algorithm illustrated in Chapter 5.
- The calibration method illustrated in this work used flat target objects at an unknown distance and brought into focus using the TTF. The accuracy and the ray tracing can be achieved in one calibration process if the target object distance is known with multiple target objects in the scene. This would help to refocus the captured image and calibration can be more accurate.
- The MLA-Ps assembly gap has been a key information to boost the Lytro cameras as engineering devices in practical applications. The above said calibration method would be enough and provide enough key parameters to calculate MLA-Ps gap.
- The MLA-Ps gap plays an important role in the MLA based LF cameras, hence the effect of variation in the gap size will help to understand the results of the cameras. Since the raw image formed on the photosensor entirely depends on the MLA placement and the ray tracing feature necessary for depth measurement will be affected if the gap is higher, leading to optical cross-talk between lenslets. Hence this work suggests the need to evaluate the reduction of the MLA-Ps gap less than the focal length of the corresponding MLA.
- Once the key parameters are calculated by above said method or by any other means, the ray projection method of depth calculation can be used to generate the absolute depth map details.

# APPENDIX





## 10 Appendix

### 10.1. Lytro-I generation camera Specifications


#### Camera: Lytro-I generation



- **Lens**
  - Focal Length: 43mm-344mm
  - Zoom: 8x optical
  - Aperture: Constant f/2.0
- **Image Sensor**
  - Sensor Type: CMOS
  - Light Field Resolution: 11 Megaray (the number of light rays captured by the light field sensor)
  - Active Area: 4.6mm x 4.6mm
- **Image**
  - Format: Light Field Picture (.lfp)
  - Aspect Ratio: 1:1
  - 2D Export Resolution: 1080 x 1080 pixels (approx. 1MP peak output)
  - File/Picture Storage: 750 - 16GB Model, 350 - 8GB Model (on-camera), additional free storage on Lytro Web.
- **Exposure**
  - Modes: Full Auto, Full Manual, Shutter Priority or ISO Priority
  - Shutter Priority: 1/250 – 8 seconds
  - ISO Priority: 80-3200
  - Exposure Lock: Yes
  - ND Filter: 4-Stop
  - Control Interface: Tap on the touchscreen
- **Screen**
  - Touchscreen: Yes
  - Size: 1.52” (diagonal)
  - Screen Type: Back-lit LCD
  - Live View: Yes
- **Playback**
  - In-camera Picture Review: Yes
- **Power**
  - Battery: Built-in, rechargeable long-life lithium-ion
  - Battery Charging: Via Micro-USB to a computer or LYTRO Fast Charger
- **External**
  - Controls: Power Button, Shutter Button, Zoom Slider, Touchscreen
  - USB: Micro-USB
  - Tripod Socket: Available via Lytro custom accessory mount, sold

- separately
- **Miscellaneous**
    - Software: Includes LYTRO Desktop for importing, organizing, processing, and interacting with living pictures. See more on the LYTRO Desktop Fact Sheet
    - Wireless: 802.11b/g/n, Wi-Fi Protected Access (WPA2)
    - E-Waste: RoHS Certified
  - **Materials**
    - Lightweight anodized aluminium with silicone grip Camera Kit Includes: LYTRO Magnetic Lens Cap, Lens cleaning cloth, wrist strap, USB cable for data transfer and charging and Quick Start guide
  - **Dimensions/Weight (Camera)**
    - 41mm x 41mm x 112mm (1.6" x 1.6" x 4.4")  
214 g (7.6 oz)

**10.2. Lytro Illum camera specifications**

<b>Camera: Lytro Illum</b>	
	
Product	LYTRO ILLUM
Dimensions	3.4"x5.7"x6.5" (86 mm x 145 mm x 166 mm)
Weight	940 grams / 33.15 oz / 2.07 lbs
Body	Magnesium and Aluminium
Grip and lens rings	Silicone
<b>Lens</b>	
Focal Length (35 mm equivalent)	9.5 - 77.8 mm (30 - 250 mm equivalent)
Crop Factor	3.19
Zoom	8x
Lens Aperture	Constant f/2.0
Macro	Focus to 0 mm from lens front
Macro Ratio	1:03

<b>Image Sensor</b>	
Sensor Type	CMOS
Light Field Resolution	40 Megaray
Processor	Snapdragon® processor by QUALCOMM® Incorporated
Sensor Format	1/1.2"
Active Area	10.82 x 7.52 mm
<b>Image</b>	
Format	Light Field Picture
2D export resolution	4MP peak output
Custom White Balance	Yes
File/Picture Storage	SD memory card slot (SD card not included)
<b>Shutter</b>	
Shutter Type	Focal plane
Fastest Shutter Speed	1/4000 sec
Slowest Shutter Speed	32 Sec
Continuous Shooting Options	Single or Continuous
Self-Timer	Yes
<b>Exposure</b>	
Exposure Metering System	Scene Evaluative
Exposure Histogram	In Live View and Playback
Exposure Modes	Program, ISO Priority, Shutter Priority, Manual
Exposure Compensation	Yes
Exposure Bracketing	Yes
Exposure Lock	Yes
Clipping Warning	Yes
<b>Focus/Autofocus</b>	

Auto-focus Modes	Region AF
<b>Screen</b>	
Touchscreen	Yes
Screen Size	4" LCD rear screen
Screen Resolution	480 x 800
Screen Angle of View	Up to 80 degrees
Adjustable Brightness	Yes
Screen Type	back-lit LCD
Articulated Angles	-10 to +90
Articulated LCD	Dual hinge tilting
Live view	Yes
<b>Playback</b>	
In-Camera Picture Review	Yes
Light Field Playback function	Refocus
<b>Menus/Interface</b>	
Customizable Buttons	Yes
<b>Power</b>	
Battery	Removable Li-Ion battery
Battery Charging	Standalone wall charger and USB
<b>External</b>	
Hot-shoe	ISO compatible hot shoe with centre pin sync manual and Lytro-TTL
Tripod Socket	Standard 1/4"-20
Cable Shutter Release Compatible	Yes
USB	Micro USB 3.0
<b>Miscellaneous</b>	

Technology	Lytro Light Field Sensor and Lytro Light Field Engine 2.0
Wireless	802.11a/b/g/n/ac enabled
Software	Includes a free desktop application for importing, processing and interacting with living pictures from the camera. The software requires Mac OS 10.8.5 or higher or Windows 7 or 64-bit Windows 8.

### 10.3. Disparity estimation

Disparity estimation using global energy function,  $E$ , for disparity image,  $D$  [110].

$$E(D) = \sum_p \left( c(p, D_p) + \sum_{q=1}^{N_p} p_1 I(|D_p - D_q|=1) + \sum_{q=1}^{N_p} p_2 I(|D_p - D_q|>1) \right)$$

where

$E(D)$  is the energy for disparity image( $D$ )

$p, q$  represent indices for pixels in the image (vectorised form)

$N_p$  is the neighborhood of the pixel  $p$  (limited to radius of the Corresponding lenslets, 7 for the Lytro-I and 13 for the Illum cameras )

$C(p, D_p)$  is the cost of pixel matching with disparity in  $D$  the  $p$

$p_1$  is the penalty passed by the user for a change in disparity values of 1 between neighbouring pixels

$p_2$  is the penalty passed by the user for a change in disparity values greater than 1 between neighboring pixels

$I[.]$  is the function which returns 1 if the argument is true and 0 otherwise

The minimized function produces a perfect disparity map with smoothing governed by parameters  $p_1$  and  $p_2$ ; however, minimizing the function for a 2D image space is a nondeterministic polynomial time problem. The semi-global matching function approximates the 2D minimization by performing multiple 1D, or linear, minimizations. The matching function aggregates costs on multiple paths which converge on the pixel under examination. Cost is computed for the disparity range specified by the minimum

disparity and number of disparities parameters. By default, In this research, the matching algorithm aggregates costs for 8 directions.

The sum of all cost for pixel  $p$  and disparity  $d$  is given by

$$S(p, d) = \sum_r L_r(p, d)$$

where  $r$  is a direction used for converging to the pixel  $p$  and  $L_r(p, d)$  is the minimum cost of the path taken in direction  $r$  from pixel

The total cost  $L_r(p, d)$  is given in the following equation:

$$L_r(p, d) = C(p, d) + \min(L_r(p - r, d), L_r(p - r, d - 1) + P_1, L_r(p - r, d + 1) + P_1, \min|L_r(p - r, i)| + P_2) - \min|p - r, k|$$

The equation uses the following costs to find the disparity by adding current cost,  $C(p, d)$  to previous pixel in direction  $r$ :

- The minimum of the cost at previous pixel with disparity  $d$
- The cost at previous pixel with disparity  $d - 1$  and  $d + 1$  with added penalty  $P_1$
- The cost at previous pixel with disparities less than  $d - 1$  and greater than  $d + 1$  with added penalty  $P_2$

In order to limit the ever-increasing value of  $L_r(p, d)$  on the path, minimum value of the previous pixel is subtracted. The upper value of  $L_r(p, d)$  is bounded by  $C_{\max} + P_2$ , where  $C_{\max}$  is the maximum value of cost  $C$ . The cost function  $C(p, d)$  is computed by

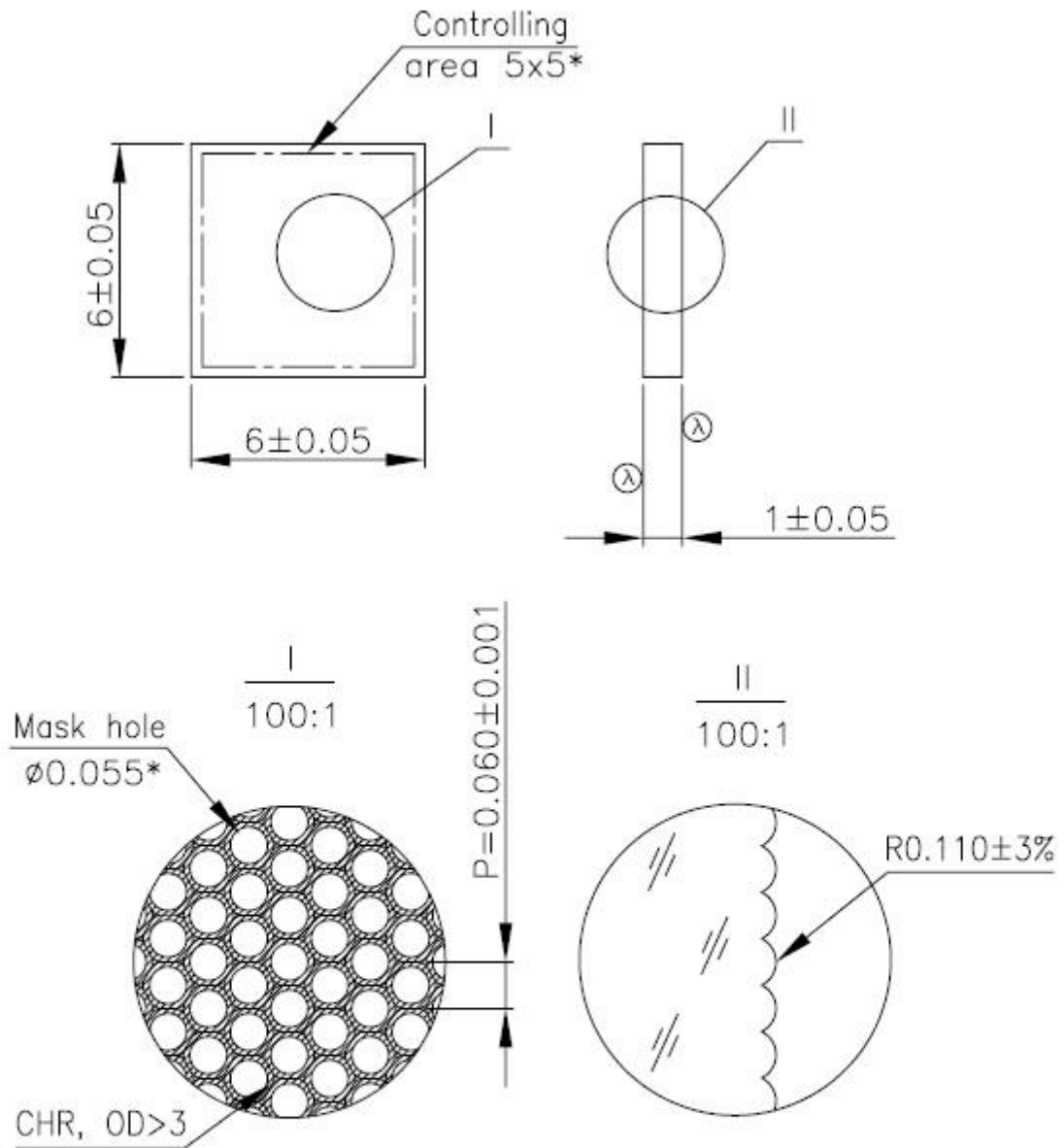
$$C(p, d) = \min(d(p, p - d, I_L, I_R), d(p - d, p, I_R, I_L))$$

where  $I_L, I_R$  are left and right sub-aperture images, respectively

$$d(p - p - d, I_L, I_R) = \min|I_L(p) - I_L(q)|$$

The value of  $C$  is aggregated over a window of a user-defined size. After computing  $S(p, d)$  for each pixel  $p$  for each disparity  $d$ , the algorithm chooses the disparity which provides the minimum cost for that pixel.

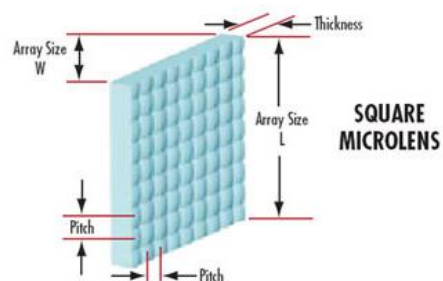
## 10.4. Thorlabs MLA Specification Sheet



1. \* Informative values
2. Focal length tolerance:  $F = 0.240 \text{ mm} \pm 3\% @ 633 \text{ nm}$
3. Non-toleranced data: ISO 10110-11
4. Defects outside the controlling area are not defined
5.  $\text{\textcircled{\scriptsize A}}$  AR-coating  $R < 1\% @ 350 \dots 950 \text{ nm}$

## 10.5. Edmund Optics MLA Specification Sheet

Coating:	Uncoated
Substrate:	Fused Silica
Radius R (mm):	8.60
Thickness (mm):	1.20
Type:	Lens Array
Pitch ( $\mu\text{m}$ ):	300
Dimensional Tolerance (mm):	$\pm 0.05$
Wavelength Range (nm):	200 - 2200
Effective Focal Length EFL (mm):	18.8
Type of Optics:	Plano-Convex
Dimensions (mm):	10.0 x 10.0
Thickness Tolerance (mm):	$\pm 0.05$
Divergence Angle ( $^\circ$ ):	$\pm 0.5$
Lens Profile:	Spherical
Pitch Tolerance ( $\mu\text{m}$ ):	$\pm 0.25$
RoHS:	Compliant





## 10.6. Homography Matrix

We estimate the homography ( $h$ ) between two views by first extracting point correspondences (control points) and then using the relationship between the planar homography and the epipolar geometry of the scene. A RANSAC algorithm is then used to remove outliers from the set of point correspondences [139]. In homogeneous coordinate system, consider two corresponding points  $x$  and  $x'$ . Let  $c$  be a non-zero constant,  $(u \ v \ 1)^T$  represents  $x'$ ,  $(x \ y \ 1)^T$  represents  $x$ , and  $H$  is given by

$$c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}$$

This equation can be modified into

$$-h_1x - h_2y - h_3 + (h_7x + h_8y + h_9)u=0$$

$$-h_4x - h_5y - h_6 + (h_7x + h_8y + h_9)v=0,$$

which can be written as

$$A_i h = 0$$

Hence least square method is used to find the matrix  $H$ .

## REFERENCE



## 11 Reference

- [1] W. Kim *et al.*, “A 1.5Mpixel RGBZ CMOS Image Sensor for Simultaneous Color and Range Image Capture,” in *IEEE International Solid-State Circuits Conference*, 2012, pp. 392–394.
- [2] J. W. Roach and J. K. Aggarwal, “Determining the Movement of Objects from a Sequence of Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 6, pp. 554–562, 1980.
- [3] T. D. Williams, “Depth from Camera Motion in a Real World Scene,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 6, pp. 511–516, 1980.
- [4] N. Ramakant, “Depth Measurement by Motion Stereo,” *Computer Graphics and Image Processing*, vol. 5, pp. 203–214, 1976.
- [5] A. Pentland, T. Darrell, M. Turk, and W. Huang, “A simple, Real-Time Range Camera,” in *Proceedings CVPR '89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1989, pp. 256–261.
- [6] A. P. Pentland, “A New Sense for Depth of Field,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 4, pp. 523–531, Jul. 1987.
- [7] R. A. Jarvis, “A Perspective on Range Finding Techniques for Computer Vision,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 2, pp. 122–139, Mar. 1983.
- [8] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo, “Operating Principles of Structured Light Depth Cameras,” in *Time-of-Flight and Structured Light Depth Cameras*, 2016, pp. 43–79.
- [9] T. Jia, Z. Zhou, and H. Gao, “Depth Measurement Based on Infrared Coded Structured Light,” *Journal of Sensors*, vol. 2014, pp. 1–8, Oct. 2014.
- [10] B. Carrihill and R. Hummel, “Experiments with the Intensity Ratio Depth Sensor,” *Computer Vision, Graphics, and Image Processing*, vol. 32, no. 3, pp. 337–358, Dec. 1985.

- [11] A. W. Lohmann and G. P. Weigelt, "The Measurement of Depth Motion by Speckle Photography," *Optics Communications*, vol. 17, no. 1, pp. 47–51, Apr. 1976.
- [12] I. U. Perwass, C. Perwass, K. De, U. Perwass, and C. Perwass, "Digital Imaging System, Plenoptic Optical Device and Image Data Processing Method," *US Patent 8,619,177*, vol. 2, no. 12, 2013.
- [13] Y. Ng and Yi-Ren Ng, "Interactive Refocusing of Electronic Images," *US 8,559,705 B2*, 2013.
- [14] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2, pp. 673–680, 2013.
- [15] J. Diebel and S. Thrun, "An Application of Markov Random Fields to Range Sensing," *Advances in Neural Information Processing Systems*, 2006.
- [16] E. Hecht, *Optics*, 10th ed. 2017.
- [17] C. Zhang and T. Chen, "Light Field Sampling," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–102, 2006.
- [18] E. Y. Lam, "Computational Photography with Plenoptic Camera and Light Field Capture: Tutorial," *Journal of the Optical Society of America A*, vol. 32, no. 11, pp. 2021–2032, 2015.
- [19] E. H. Adelson and J. R. Bergen, "The Plenoptic Function and the Elements of Early Vision," MIT Press, 1991.
- [20] M. Levoy and P. Hanrahan, "Light Field Rendering," *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques - SIGGRAPH '96*, pp. 31–42, 1996.
- [21] R. Ng, M. Levoy, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-held Plenoptic Camera," *Informational*, pp. 1–11, 2005.
- [22] T. Fujii and M. Tanimoto, "Free-Viewpoint TV System Based on Ray-Space Representation," *SPIE ITCOM*, vol. 4864, pp. 175–189, Nov. 2002.

- [23] H. H. Barrett, K. J. Myers, and S. Rathee, *Foundations of Image Science*, vol. 31, no. 4. Wiley-Interscience, 2004.
- [24] R. Ramamoorthi and P. Hanrahan, “On the Relationship Between Radiance and Irradiance: Determining the Illumination from Images of a Convex Lambertian Object,” *Journal of the Optical Society of America A*, vol. 18, no. 10, pp. 2448–2459, 2001.
- [25] *Shorter Oxford English Dictionary*. 1968.
- [26] S. Prabu and M. S. Moorthi, “Depth Reconstruction using Geometric Correction with Anaglyph Approach for Satellite Imagery,” *International Conf. on Advances in Communication, Network, and Computing*, 2014.
- [27] F. E. Ives, “Parallax Stereogram and Process of Making Same,” 1903.
- [28] T. Georgiev, “100 Years Light-Field Directional Imaging,” pp. 1–31, 2008.
- [29] K. Takahashi and T. Naemura, “Layered Lightfield Rendering with Focus Measurement,” *Signal Processing: Image Communication*, vol. 21, no. 6, pp. 519–530, Jul. 2006.
- [30] R. Ng, “Digital light field photography,” Stanford University, 2006.
- [31] H. Ives, “Parallax Panoramagrams Made With a Large Diameter Lens,” *Journal of the Optical Society of America*, vol. 20, no. 6, pp. 332–340, Jun. 1930.
- [32] H. E. Ives, “Optical Properties of a Lippmann Lenticulated Sheet,” *Journal of the Optical Society of America*, vol. 21, no. 3, pp. 171–176, Mar. 1931.
- [33] 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,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3280–3287, 2013.
- [34] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, “Using Plane + Parallax for Calibrating Dense Camera Arrays,” *Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 2–9, 2004.
- [35] A. Isaksen, L. Mcmillan, and S. J. Gortler, “Dynamically Reparameterized Lightfields,” *Annual Conference on Computer Graphics and Interactive*

*Techniques - SIGGRAPH*, pp. 297–306, 2000.

- [36] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, “Light Field Microscopy,” *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 924–934, 2006.
- [37] M. Broxton *et al.*, “Wave Optics Theory and 3-D Deconvolution for the Light Field Microscope,” *Optics Express*, vol. 21, no. 21, pp. 25418–25439, Oct. 2013.
- [38] T. Georgiev and A. Lumsdaine, “Focused Plenoptic Camera and Rendering,” *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 1–11, 2010.
- [39] T. Georgiev, K. Zheng, and B. Curless, “Spatio-Angular Resolution Tradeoffs in Integral Photography,” *Eurographics Symposium on Rendering (EGSR)*, pp. 263–272, 2006.
- [40] C. Wu, M. McCormick, A. Aggoun, and S. Y. Kung, “Depth Mapping of Integral Images Through Viewpoint Image Extraction With a Hybrid Disparity Analysis Algorithm,” *Journal of Display Technology*, vol. 4, no. 1, pp. 101–108, Mar. 2008.
- [41] T. E. Bishop, S. Zanetti, and P. Favaro, “Light Field Superresolution,” *IEEE International Conference on Computational Photography, ICCP*, 2009.
- [42] C. Perwaß, L. Wietzke, and R. Gmbh, “Single Lens 3D-Camera with Extended Depth-of-Field,” *Proceedings of SPIE*, vol. 8291/108, no. 431, pp. 1–15, 2012.
- [43] T. Georgiev, A. Lumsdaine, and S. Goma, “Plenoptic Principal Planes,” in *Imaging and Applied Optics*, 2011, no. 2, p. JTuD3.
- [44] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, “Image and Depth from a Conventional Camera with a Coded Aperture,” *ACM Transactions on Graphics*, vol. 26, no. 3, p. 70:1-9, 2007.
- [45] N. Li, H. Lin, B. Sun, M. Zhou, and J. Yu, “Rotational Crossed-Slit Light Fields,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4405–4413.
- [46] “Light Fields and Computational Photography.” [Online]. Available: <https://graphics.stanford.edu/projects/lightfield/>. [Accessed: 21-Sep-2015].
- [47] E. H. Adelson and J. Y. a Wang, “Single Lens Stereo with a Plenoptic Camera,”

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 99–106, 1992.
- [48] C.-K. Liang *et al.*, “Programmable Aperture Photography,” in *ACM SIGGRAPH 2008 papers on - SIGGRAPH '08*, 2008, vol. 27, no. 3, p. 55:1-10.
- [49] M. Z. Alam and B. K. Gunturk, “Hybrid light field imaging for improved spatial resolution and depth range,” *Machine Vision and Applications*, pp. 1–12, Nov. 2017.
- [50] T. G. Georgiev, A. Lumsdaine, and S. Goma, “High Dynamic Range Image Capture with Plenoptic 2.0 Camera,” in *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest*, 2009, pp. 1–3.
- [51] Lytro Inc, “Lytro - Home,” 2016. [Online]. Available: <https://www.lytro.com/>. [Accessed: 23-Feb-2015].
- [52] “Raytrix | 3D light field camera technology.” [Online]. Available: <http://www.raytrix.de/>. [Accessed: 24-Sep-2015].
- [53] B. Newhall, *The History of Photography from 1839 to the Present Day; Book*. The Museum of Modern Art.; New York, 1964.
- [54] Alasia Alfred Victor, “Three-dimensional camera,” US3524395 A, 1969.
- [55] J. P. Mellor and S. Teller, “Dense Depth Maps from Epipolar Images,” 1996.
- [56] G. I. Giora Yahav, “Three dimensional camera,” US6100517 A, 1996.
- [57] M. Subbarao, “Parallel Depth Recovery by Changing Camera Parameters,” in *Second International Conference on Computer Vision*, 1988, pp. 149–155.
- [58] B. Barsky and T. J. Kosloff, “Algorithms for Rendering Depth of Field Effects in Computer Graphics,” *World Scientific and Engineering Academy and Society (WSEAS)*, pp. 999–1010, 2008.
- [59] H. Supèr and A. Romeo, “Coding Depth Perception from Image Defocus,” vol. 105, pp. 199–203, 2014.
- [60] Y. Y. Schechner and N. Kiryati, “Depth from Defocus vs. Stereo: How Different Really are they?,” *International Journal of Computer Vision*, vol. 39, no. 2, pp.

141–162, 2000.

- [61] C. Zhou, S. Lin, and S. K. Nayar, “Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring,” *International Journal of Computer Vision*, vol. 93, no. 1, pp. 53–72, 2011.
- [62] J. Ens and P. Lawrence, “An Investigation of Methods for Determining Depth from Focus,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 97–108, 1993.
- [63] A. Lumsdaine and T. Georgiev, “Full resolution lightfield rendering,” *Indiana University and Adobe Systems, Tech. Rep*, no. January, pp. 1–12, 2008.
- [64] Y. Luan, X. He, B. Xu, P. Yang, and G. Tang, “Automatic Calibration Method for Plenoptic Camera,” *Optical Engineering*, vol. 55, no. 4, p. 043111:1-5, Apr. 2016.
- [65] “The Advantages of Telecentricity.” [Online]. Available: <https://www.edmundoptics.eu/resources/application-notes/imaging/advantages-of-telecentricity/>. [Accessed: 29-Dec-2016].
- [66] D. Weimer, H. Thamer, C. Fellmann, M. Lütjen, K.-D. D. Thoben, and B. Scholz-Reiter, “Towards 100% In-Situ 2D/3D Quality Inspection Of Metallic Micro Components Using Plenoptic Cameras,” *CIRP Conference on Manufacturing Systems*, vol. 17, pp. 847–852, 2014.
- [67] X. Guo, H. Lin, Z. Yu, and S. McCloskey, “Barcode Imaging using a Light Field Camera,” *ECCV Workshops*, vol. 2, pp. 519–532, 2014.
- [68] M. G. C. T. R. K. A. Kolb, “Time-of-Flight and Depth Imaging,” *Springer*, vol. 24, no. 2, p. iii, 2010.
- [69] G. Surya and M. Subbarao, “Depth from Defocus by Changing Camera Aperture: A Spatial Domain Approach,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1993)*, pp. 61–67, 1993.
- [70] D. G. Dansereau, O. Pizarro, and S. B. Williams, “Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.



1027–1034, 2013.

- [71] A. S. Gujrathi and D. Gehlot, “Testing and Performance of the Convex Lens Concentrating Solar Power Panel Prototype,” *International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com ISO Certified Journal*, vol. 9001, no. 6, 2250.
- [72] M. C. Hutley and M. Arrays, “Microlens arrays,” *Physics Education*, vol. 29, no. 3, pp. 160–163, 1999.
- [73] J. Aizenberg and G. Hendler, “Designing Efficient Microlens Arrays: Lessons from Nature,” *Journal of Materials Chemistry*, vol. 14, no. 14, p. 2066, Jul. 2004.
- [74] P. Nussbaum, R. Völkel, and H. P. Herzig, “Design, Fabrication and Testing of Microlens Arrays for Sensors and Microsystems,” *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3099, pp. 196–211, 1997.
- [75] J. Chang, I. Kauvar, X. Hu, and G. Wetzstein, “Variable Aperture Light Field Photography: Overcoming the Diffraction-Limited Spatio-Angular Resolution Tradeoff,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3737–3745.
- [76] M. Sieler, P. Schreiber, and A. Bräuer, “Microlens Array Based LCD Projection Display with Software-only Focal Distance Control,” *Proceedings of SPIE*, vol. 8643, p. 86430B, Mar. 2013.
- [77] M. K. Hedili, M. O. Freeman, and H. Urey, “Microlens Array-Based High-Gain Screen Design for Direct Projection Head-Up Displays.,” *Applied optics*, vol. 52, no. 6, pp. 1351–7, Feb. 2013.
- [78] J.-W. Pan, C.-M. Wang, H.-C. Lan, W.-S. Sun, and J.-Y. Chang, “Homogenized LED-Illumination Using Microlens Arrays for a Pocket-Sized Projector.,” *Optics express*, vol. 15, no. 17, pp. 10483–10491, 2007.
- [79] S. E. Jiangeng Xue, Jason David Myers, “Thin Film Photovoltaic Devices with Microlens Arrays,” 2012.
- [80] W. Sweatt *et al.*, “Micro-Optics for High-Efficiency Optical Performance and

- Simplified Tracking for Concentrated Photovoltaics (CPV),” *International Optical Design Conference and Optical Fabrication and Testing*, p. ITuC4, 2010.
- [81] C. Birklbauer and O. Bimber, “Panorama Light-Field Imaging,” 2014.
- [82] J. Mertz, *Introduction to optical microscopy*. Roberts, 2010.
- [83] C. Hapter *et al.*, “Fundamental Optical Design,” in *Caos & Tka*, vol. 1997, no. 1, 2002, pp. 183–197.
- [84] N. Sabater, M. Seifi, V. Drazic, G. Sandri, and P. Pérez, “Accurate disparity estimation for plenoptic images,” in *Lecture Notes in Computer Science*, 2015, vol. 8926, pp. 548–560.
- [85] Y. Bok, H. G. Jeon, and I. S. Kweon, “Geometric Calibration of Micro-Lens-Based Light Field Cameras Using Line Features,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 287–300, 2017.
- [86] Y. Zheng *et al.*, “Single-Image Vignetting Correction,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 12, pp. 2243–56, 2009.
- [87] H. J. Jaesik, P. Gyeongmin, and C. Jinsun, “Accurate Depth Map Estimation from a Lenslet Light Field Camera,” *Cvpr*, 2015.
- [88] C. M. Thomason, B. S. B. Thurow, T. T. W. Fahringer, and B. S. B. Thurow, “Calibration of the Microlens Array of a Plenoptic Camera,” *52nd Aerospace Sciences Meeting*, no. January, pp. 1–18, 2014.
- [89] “Lytro Desktop.” [Online]. Available: <https://illum.lytro.com/desktop>. [Accessed: 01-Jun-2015].
- [90] O. Johannsen, C. Heinze, B. Goldluecke, and C. Perwaß, “On the calibration of focused plenoptic cameras,” 2013, vol. 8200 LNCS, pp. 302–317.
- [91] C. Meah *et al.*, “Towards Plenoptic Multi-View Imaging,” *Imaging and Applied Optics 2015*, vol. 42, p. JT5A.42, 2015.
- [92] R. Marshall, I. Styles, E. Claridge, and K. Bongs, “Plenoptic Imaging of the Retina: Can it Resolve Depth in Scattering Tissues?,” *Biomedical Optics 2014*, vol. 604460, no. 170, p. BM3A.60, 2014.

- [93] R. J. Marshall *et al.*, “Optimisation of Patterned Illumination to Improve Depth Estimation From a Plenoptic Camera,” *Imaging and Applied Optics 2015*, vol. 41, p. JT5A.41, 2015.
- [94] P. Fankhauser, M. Bloesch, D. Rodriguez, R. Kaestner, M. Hutter, and R. Siegwart, “Kinect v2 for Mobile Robot Navigation: Evaluation and Modeling,” 2015.
- [95] S. Rangappa, M. Taylor, J. Petzing, P. Kinnell, and M. Jackson, “The suitability of Lightfield Camera Depth Maps for Coordinate Measurement Applications,” *Eighth International Conference on Machine Vision (ICMV 2015)*, vol. 9875, p. 987523, Dec. 2015.
- [96] S. Piano, D. Sims-Waterhouse, and R. K. Leach, “Fusion of light-field and photogrammetric surface form data,” in *Applied Optical Metrology II*, 2017, vol. 10373, p. 4.
- [97] “Photography - black-and-white pictorial still camera negative film/process systems - determination of ISO speed,” 1993.
- [98] “Distortion (optics).” [Online]. Available: [https://en.wikipedia.org/wiki/Distortion\\_\(optics\)](https://en.wikipedia.org/wiki/Distortion_(optics)). [Accessed: 21-Jul-2015].
- [99] G. Sibley, *Long Range Stereo Data-Fusion From Moving Platforms*. 2007.
- [100] P. Pinggera, U. Franke, and R. Mester, “High-Performance LongRange Obstacle Detection Using Stereo Vision,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 1308–1313.
- [101] A. C. She, K. D. Hjelmstad, and T. S. Huang, “Nondestructive Evaluation of Civil Structures and Materials Using Stereo Camera Measurements,” in *[1992] Proceedings. 11th IAPR International Conference on Pattern Recognition*, pp. 708–711.
- [102] X. Ding, L. Xu, H. Wang, X. Wang, and G. Lv, “Stereo Depth Estimation Under Different Camera Calibration and Alignment Errors,” *Applied Optics*, vol. 50, no. 10, p. 1289, Apr. 2011.
- [103] F. Pérez, A. Pérez, M. Rodríguez, and E. Magdaleno, “A Fast and Memory-

- Efficient Discrete Focal Stack Transform for Plenoptic Sensors,” *Digital Signal Processing*, vol. 38, pp. 95–105, Mar. 2015.
- [104] H. Lin, C. Chen, S. B. Kang, and J. Yu, “Depth Recovery from Light Field Using Focal Stack Symmetry,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, vol. 2015 Inter, pp. 3451–3459.
- [105] M. J. Kim, T. H. Oh, and I. S. Kweon, “Cost-Aware Depth Map Estimation for Lytro Camera,” in *2014 IEEE International Conference on Image Processing, ICIP 2014*, 2014, pp. 36–40.
- [106] C. Hahne, A. Aggoun, S. Haxha, V. Velisavljevic, J. C. J. Fernández, and J. C. J. Fern, “Baseline Of Virtual Cameras Acquired by a Standard Plenoptic Camera Setup,” *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, pp. 1–3, 2014.
- [107] D. G. Dansereau, O. Pizarro, and S. B. Williams, “Linear Volumetric Focus for Light Field Cameras,” *ACM Transactions on Graphics*, vol. 34, no. 2, pp. 1–20, 2015.
- [108] D. Lelescu and F. Bossen, “Representation and Coding of Light Field Data,” *Graphical Models*, vol. 66, no. 4, pp. 203–225, Jul. 2004.
- [109] C. Kolb, D. Mitchell, and P. Hanrahan, “A Realistic Camera Model for ComputerGraphics,” in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*, 1995, pp. 317–324.
- [110] D. A. Kerr, “Principle of the Split Image Focusing Aid and the Phase Comparison Autofocus Detector in Single Lens Reflex Cameras,” 2005.
- [111] W. Sun, X. Yang, S. Xiao, and W. Hu, “Robust Checkerboard Recognition for Efficient Nonplanar Geometry Registration in Projector-camera Systems,” *PROCAMS*, 2008.
- [112] A. Geiger, F. Moosmann, O. Car, and B. Schuster, “Automatic Camera and Range Sensor Calibration Using a Single Shot.,” in *IEEE Conference on Robotics and Automation (ICRA)*, 2012, pp. 3936–3943.
- [113] P. Yang *et al.*, “Close-Range Photogrammetry with Light Field Camera: From

- Disparity Map to Absolute Distance,” *Applied Optics*, vol. 55, no. 27, p. 7477, Sep. 2016.
- [114] P. Harvey, “ExifTool: Read, Write and Edit Meta Information,” 2013. .
- [115] H. Hirschmuller, “Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information,” *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Volume: 2)*, vol. 2, no. 2, pp. 807–814, 2005.
- [116] K. Tarabanis, R. Y. Tsai, and D. S. Goodman, “Calibration of a Computer Controlled Robotic Vision Sensor with a Zoom Lens,” *Computer Vision, Graphics, and Image Processing*, vol. 59, no. 2, pp. 226–241, Mar. 1994.
- [117] Z. Zhang, “Flexible Camera Calibration By Viewing a Plane From Unknown Orientations.”
- [118] S. S. Beauchemin, R. Bajcsy, and G. Givaty, “Modelling and Removing Radial and Tangential Distoritions in Spherical Lenses.”
- [119] J. Jedlička and M. Potůčková, “Correction of Radial Distortion in Digital Images,” *Proceedings Technical Computing Prague, 2007*.
- [120] M. D. Grossberg and S. K. Nayar, “The Raxel Imaging Model and Ray-Based Calibration,” *International Journal of Computer Vision*, vol. 61, no. 2, pp. 119–137, 2005.
- [121] J.-Y. Rau and P.-C. Yeh, “A Semi-Automatic Image-Based Close Range 3D Modeling Pipeline Using a Multi-Camera Configuration,” *Sensors*, vol. 12, no. 12, pp. 11271–11293, Aug. 2012.
- [122] Y. Xu, K. Maeno, H. Nagahara, R.-I. ichiro Taniguchi, H. Magahara, and R.-I. ichiro Taniguchi, “Camera Array Calibration for Light Field Acquisition,” *Front. Comput. Sci*, vol. 9, no. 5, pp. 691–702, 2015.
- [123] G. D’Emilia and D. Di Gasbarro, “Review of Techniques for 2D Camera Calibration Suitable for Industrial Vision Systems,” in *Journal of Physics: Conference Series*, 2017, vol. 841, no. 1, pp. 2–5.
- [124] B. Louren, “Depth Estimation using Light-Field Cameras,” 2014.

- [125] C. Heinze, “Design and Test of a Calibration Method for the Calculation of Metrical Range Values for 3D Light Field Cameras,” Hamburg University of Applied Sciences, 2014.
- [126] Z. Zhang, “A Flexible New Technique for Camera Calibration,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22.11, pp. 1330–1334, 2000.
- [127] Robert J. Burger, “Lenslet Array Systems and Methods,” 1997.
- [128] R. T. Held, E. a. Cooper, and M. S. Banks, “Blur and Disparity are Complementary Cues to Depth,” *Current Biology*, vol. 22, no. 5, pp. 426–431, 2012.
- [129] D. G. Dansereau, “Light Field Toolbox for Matlab,” 2013.
- [130] T. E. Bishop and P. Favaro, “The light field camera: Extended depth of field, aliasing, and superresolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 972–986, 2012.
- [131] H. Sun, “Thin Lens Equation for a Real Laser Beam with Weak Lens Aperture Truncation,” *Optical Engineering*, vol. 37, no. 11, p. 2906, Nov. 1998.
- [132] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens, and D. Vandermeulen, “Feature Detection on 3D Face Surfaces for Pose Normalisation and Recognition,” in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2010, pp. 1–6.
- [133] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An Efficient Alternative to SIFT or SURF,” in *2011 International Conference on Computer Vision*, 2011, pp. 2564–2571.
- [134] N. Zeller, C. A. Noury, F. Quint, C. Teulière, U. Stilla, and M. Dhome, “Metric Calibration of a Focused Plenoptic Camera Based on a 3D Calibration Target,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 449–456, 2016.
- [135] R. Danzl, F. Helmlí, and S. Scherer, “Focus Variation – A New Technology for High Resolution Optical 3D Surface Metrology,” *10th International Conference*

*of Slovenian Society for Non-Destructive Testing*, pp. 1–10, 2009.

- [136] R. Danzl, F. Helml, and S. Scherer, “Focus Variation – A Robust Technology for High Resolution Optical 3D Surface Metrology,” *Strojniški vestnik – Journal of Mechanical Engineering*, vol. 2011, no. 03, pp. 245–256, Mar. 2011.
- [137] C. Hahne, A. Aggoun, and V. Velisavljevic, “Refocusing Distance of a Standard Plenoptic Camera-I,” *3DTV-Conference*, vol. 2015–July, 2015.
- [138] A. M. A. R. A. Ggoun, C. H. H. Ahne, V. L. V Elisavljevic, S. U. F. Iebig, M. A. P. Esch, and U. Kingdom, “Refocusing Distance of a Standard Plenoptic Camera,” *Optical Society of America*, vol. 24, no. 19, pp. 21521–21540, 2016.
- [139] K. Mueller, T. Moller, and R. Crawlis, “Splating without the blur,” *Proceedings Visualization '99 (Cat. No.99CB37067)*, 1999.