Title: Hybrid light field image super-resolution and interpolation method using multi-array cameras

Name: Dr. Saber Ahmed Ahmed Farag

This dissertation is a digitised version submitted to the University of Bedfordshire.

It is available only to view.

This Item is subject to copyright
Hybrid Light Field Image Super-resolution and Interpolation method Using Multi-array Cameras

by

SABER FARAG

UNIVERSITY OF BEDFORDSHIRE
Department of Computer Science and Technology

A thesis submitted in accordance with the requirements for the degree of

DOCTOR OF PHILOSOPHY

in the Institute for Research in Applicable Computing (IRAC)

November 2018
Declaration

Hereby, I declare that this work has been composed by myself and the work has been carried out in line with the conditions of Research Degree Regulations by the University. I also confirm that this thesis is not submitted for any other degree or professional qualification. In addition, I confirm that the thesis submitted is my own, except indicated by specific reference in the text. Research done in collaboration with, or with the assistance of, others, is indicated as such. Any observations stated in this thesis are those of the author.

SABER FARAG

2nd November 2018
Abstract

Recent advances in camera technologies have led to the design of plenoptic cameras. This camera type can capture multiple images of the same scene using arrays of microlenses, where each microlens has a shifted location providing a separate view of the scene. Such a design results in a superior performance, as compared to traditional cameras, enabling multi-view or multi-focal imaging captured in a single shot. However, the main drawback of the currently available plenoptic imaging technology is limited spatial resolution, which makes it difficult to use in applications where sharpness or high-resolution is essential, such as in the film industry. Although some previous attempts have addressed this issue, they were affected by high computational complexity as well as limited interpolation factor. To resolve this, a novel light field field hybrid super-resolution method is proposed which combines two traditional methods of multi image super-resolution and hybrid single image super-resolution to create hybrid super-resolution image for efficient application to plenoptic images. Furthermore, after this combination, the output of the hybrid super-resolution image is segmented into the objects of interest. Then, super-resolution reconstruction by sparse representation is applied to super resolve the segmented image. This technique helps to increase the resolution of light field images and maintain sharpness after super-resolution. Additionally, block matching super-resolution is proposed to provide a means of enhancement for the resolution of plenoptic images by developing corresponding super-resolution methods which exploit the disparity information, estimated from the light field images, to reduce the matching area in the super-resolution process. The proposed method is denoted as block matching super-resolution super-resolution. Following on, the proposed
A novel super-resolution method is combined with directionally adaptive image interpolation to preserve sharpness of the high-resolution images. In addition, light field digital refocusing with the proposed super-resolution approaches can be used to record the light field and provide maximum achievable resolution. With this simplification it is easy to explain the method of refocusing and the characteristics of the performance. The complexity of the standard light field camera configuration is also taken into account. The implemented super-resolution approaches that have been proposed are used to super resolve the ‘all-in-focus’ images. This research has narrowed the knowledge gap by creating a working super-resolution and interpolation application in order to allow higher quality to all light field images captured by plenoptic cameras. The significant advantage of this application for computer vision is the super-resolved micro images or the various angles available in a single light field super resolution image which allow depth estimation. Moreover, this research demonstrates a steady gain in the peak signal to noise ratio and structural similarity index quality of the super-resolved images for the resolution enhancement factor 8x8, as compared to the most recent approaches.
Dedication

To my family and friends,

To my late father, my mother, siblings and wife, who have always supported me during my study journey as well as have been praying for me during every conversation we had. Especially, my mother Fatima Ali Maqsa, who always encouraged me to be patient, despite the difficulties I have faced during my PhD. Her support taught me how to enhance my trust. As Gisele Bundchen says, “The more you trust your intuition, the more empowered you become and the happier you become”.

iii
Acknowledgments

First and foremost, in the name of ALLAH, my appreciation and gratitude go to my supervisor Dr. Vladan Velisavljevic, whose support, patience, feedback, suggestions as well as encouragement allowed my PhD life to stay constant and well organised. Special thanks are due to Prof. Amar Aggoun who was my second supervisor during the first half of my PhD journey, his advice and feedback motivated me to enhance this work.

I am grateful to Dr. Hong Qing Yu who joined the supervision team recently, his support and advice helped develop the quality of the work. Many thanks go to Dr. Shyqyri Haxah and Dr. Vladimir Dyo for their suggestions as they were internal examiners to this thesis at the two progression point stages. Also, Dr. Christopher Hahne for providing assistance with the datasets used in the thesis.

Many thanks are due to my research friends Nashat Al-masri, Sajeel Ahmed and Shehu Muhammad whose support allowed me to stay motivated during my thesis journey. Appreciation to the IRAC staff members, who provided me with all sources and requirements needed to complete the simulation part of the thesis. I would like to thank the administrator research of the RGS, Mrs Caroline Aird for her high moral, outstanding treatment and continued support. In the end, words are not enough to express my appreciation to my family, mainly my mother for providing moral support, as well as inspiring me to work hard and never lose hope whatever the circumstances. Her special advice was to continue to trust in Allah no matter the circumstances.

God says (وَمَنْ يَتَّقِي اللَّهَ يُجْعَلُ لَهُ مَرْجَعًا وَذِي رَفَعٍ مِّنْ حَيْبٍ لَا يَخْسِيَ وَمَنْ يَتَّقُونَ عَلَى اللَّهِ فَهُوَ خَشِيَّةٌ) “And unto everyone who is conscious of God, He always grants a way out of unhappiness” (Quran, Al-Talaq) صدوق الله العظيم
List of Publications


# Table of Contents

CONTENTS ........................................ PAGE

Abstract ......................................................................................................................... i
List of Abbreviations ........................................................................................................ x
List of Figures .................................................................................................................. xii
List of Tables .................................................................................................................. xv

1. Introduction .................................................................................................................. 1
   1.1 Background ............................................................................................................ 2
   1.1.1 Difference between Conventional and Plenoptic Cameras ............................ 2
   1.1.2 Concept of Super-resolution in Image Processing ........................................... 5
   1.1.3 Light Field Refocusing ..................................................................................... 7
   1.2 Limitations of Existing Work and Challenges ...................................................... 7
   1.3 Motivation of Image Super-resolution Applied to LF images .............................. 10
   1.4 Research Aims ..................................................................................................... 10
   1.5 Contributions ....................................................................................................... 11
      1.5.1 Hybrid Light Field Super-resolution ............................................................. 11
      1.5.2 Block Matching Super-resolution for Light field Images ............................. 11
      1.5.3 Light Field Image Refocusing ..................................................................... 12
   1.6 Structure of the Thesis ......................................................................................... 13

2. Literature Review ......................................................................................................... 14
   2.1 Introduction .......................................................................................................... 14
   2.2 A Brief History of Camera Technology ............................................................... 15
      2.2.1 Light Field Cameras ...................................................................................... 16
      2.2.2 Conventional Camera ................................................................................... 20
   2.3 Existing Classical Super-Resolution Methods ....................................................... 23
      2.3.1 Multi Frame Super-resolution ...................................................................... 24
      2.3.2 Example Based Super-Resolution ................................................................. 26
   2.4 Interpolation-based Methods .............................................................................. 26
      2.4.1 Non-linear interpolation methods ................................................................. 27
3. Hybrid Light Field Super-resolution Approach ................................................. 72

3.1 Introduction ................................................................................................. 73
3.2 Improving Resolution Using a Hybrid Imaging System .............................. 73
3.3 Proposed Hybrid Super-Resolution (HYSR) ............................................... 74
3.4 Development of the Novel Method for Light Field Images ......................... 80
3.4.1 Light Field Super-Resolution Approaches ............................................. 80
3.4.2 Proposed Light Field Hybrid Super-Resolution Approach ..................... 81
3.5 Evaluation Process Techniques ................................................................ 87
3.6 Tools and Data Collection ......................................................................... 89
3.7 HYSR Approach Simulation and Results .................................................. 89
References ........................................................................................................................................ 171
A. Appendix .................................................................................................................................... 193
B. Appendix .................................................................................................................................... 200
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM3D</td>
<td>Block Matching 3D</td>
</tr>
<tr>
<td>BMSR</td>
<td>Block Matching Super-resolution</td>
</tr>
<tr>
<td>DOF</td>
<td>Depth of Field</td>
</tr>
<tr>
<td>DSLR</td>
<td>Digital Single Lens Reflex</td>
</tr>
<tr>
<td>EPI</td>
<td>Epipolar Plane Image</td>
</tr>
<tr>
<td>EPIs</td>
<td>Epipolar Plane Images</td>
</tr>
<tr>
<td>FSA</td>
<td>Full Search Algorithm</td>
</tr>
<tr>
<td>GMM</td>
<td>Gaussian Mixture Model</td>
</tr>
<tr>
<td>HR</td>
<td>High Resolution</td>
</tr>
<tr>
<td>HYSISR</td>
<td>Hybrid Single Image Super-resolution</td>
</tr>
<tr>
<td>HYSR</td>
<td>Hybrid Super-resolution</td>
</tr>
<tr>
<td>KMP</td>
<td>Kernel Matching pursuit</td>
</tr>
<tr>
<td>KRR</td>
<td>Kernel Ridge Regression</td>
</tr>
<tr>
<td>LF</td>
<td>Light Field</td>
</tr>
<tr>
<td>LFBMSR</td>
<td>Light Field Block Matching Super-resolution</td>
</tr>
<tr>
<td>LFHYSR</td>
<td>Light Field Hybrid Super-resolution</td>
</tr>
<tr>
<td>LFR</td>
<td>Light Field Rendering</td>
</tr>
<tr>
<td>LFSR</td>
<td>Light Field Super-resolution</td>
</tr>
<tr>
<td>LR</td>
<td>Low Resolution</td>
</tr>
<tr>
<td>MA-XRF</td>
<td>Macro X-Ray Fluorescence</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>MISR</td>
<td>Multi Image Super-resolution</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean-Square Error</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NEDI</td>
<td>New Edge Directed Interpolation</td>
</tr>
<tr>
<td>PSF</td>
<td>Point Spared Function</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak Signal to Noise Ratio</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning Electron Microscope</td>
</tr>
<tr>
<td>SISR</td>
<td>Single Image Super-resolution</td>
</tr>
<tr>
<td>SLR</td>
<td>Single Lens Reflex</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SR</td>
<td>Super Resolution</td>
</tr>
<tr>
<td>SRR</td>
<td>Super Resolution Reconstruction</td>
</tr>
<tr>
<td>SSIM</td>
<td>Structural Similarity Index Measure</td>
</tr>
<tr>
<td>VBM4D</td>
<td>Video Block Matching 4D</td>
</tr>
<tr>
<td>VPs</td>
<td>View Point Images</td>
</tr>
<tr>
<td>WSD</td>
<td>Wiener Filter in Similarity Domain</td>
</tr>
<tr>
<td>XRF</td>
<td>X-Ray Fluorescence</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Figure 1.1</strong></td>
<td>Differences between conventional and plenoptic cameras</td>
</tr>
<tr>
<td><strong>Figure 1.2</strong></td>
<td>Raw image which captured by our plenoptic camera</td>
</tr>
<tr>
<td><strong>Figure 1.3</strong></td>
<td>The multi-frame super-resolution concept</td>
</tr>
<tr>
<td><strong>Figure 2.1</strong></td>
<td>Raw image captured by the plenoptic camera</td>
</tr>
<tr>
<td><strong>Figure 2.2</strong></td>
<td>The micro lens array has an image projected onto it by the plenoptic camera</td>
</tr>
<tr>
<td><strong>Figure 2.3</strong></td>
<td>Camera array of tiny pinholes</td>
</tr>
<tr>
<td><strong>Figure 2.4</strong></td>
<td>The Lytro camera (Ng et al., 2005)</td>
</tr>
<tr>
<td><strong>Figure 2.5</strong></td>
<td>Camera obscura (Paolucci, 2006)</td>
</tr>
<tr>
<td><strong>Figure 2.6</strong></td>
<td>Daguerreotype camera (Robinson et al., 2017)</td>
</tr>
<tr>
<td><strong>Figure 2.7</strong></td>
<td>Leica camera (Day and McNeil, 1996)</td>
</tr>
<tr>
<td><strong>Figure 2.8</strong></td>
<td>Canon camera (Phil, 2018)</td>
</tr>
<tr>
<td><strong>Figure 2.9</strong></td>
<td>Directional image interpolation edge’s flow basic processing (Sun et al., 2013)</td>
</tr>
<tr>
<td><strong>Figure 2.10</strong></td>
<td>Image interpolation locally adaptive wavelet (Chang and Cvetkovic, 2006)</td>
</tr>
<tr>
<td><strong>Figure 2.11</strong></td>
<td>Baboon magnified images</td>
</tr>
<tr>
<td><strong>Figure 2.12</strong></td>
<td>The reconstruction-based algorithm</td>
</tr>
<tr>
<td><strong>Figure 2.13</strong></td>
<td>Scheme of the hybrid example-based method</td>
</tr>
<tr>
<td><strong>Figure 2.14</strong></td>
<td>Process of downsampling images, observation model (Park et al., 2003)</td>
</tr>
<tr>
<td><strong>Figure 2.15</strong></td>
<td>Scheme for super-resolution (Alam et al., 2009)</td>
</tr>
<tr>
<td><strong>Figure 2.16</strong></td>
<td>The XRF map</td>
</tr>
<tr>
<td><strong>Figure 2.17</strong></td>
<td>Image registration before SR (Tian, 2011)</td>
</tr>
<tr>
<td><strong>Figure 2.18</strong></td>
<td>Process of image registration (Zitova abd Flusser, 2003)</td>
</tr>
<tr>
<td><strong>Figure 2.19</strong></td>
<td>Capital BM process (Cuevas et al., 2015)</td>
</tr>
</tbody>
</table>
Figure 2.20: 9 micro images .................................................................................................................................. 52
Figure 2.21: Block diagram of WSD (Gustavsson et al., 2018). ................................................................. 57
Figure 2.22: Multi refocused images in (a–e) and extended DoF in (f) (Ng et al., 2005) .................. 61
Figure 2.23: Shift-and-add algorithm (Ng et al., 2005) ............................................................................. 67
Figure 2.24: Blurred regions under-sampled shift refocus algorithm (Ng et al., 2005) .............. 68
Figure 3.1: Hybrid super-Resolution framework ......................................................................................... 75
Figure 3.2: Laplacian pyramid ..................................................................................................................... 77
Figure 3.3: Proposed hybrid super-resolution framework ................................................................. 79
Figure 3.4: Proposed hybrid light field super-resolution approach ......................................................... 82
Figure 3.5: An illustration of gauge freedom in geometry ....................................................................... 88
Figure 3.6: The Ground truth image used to test the HYSR ................................................................. 90
Figure 3.7: Comparison with SISR, HYSISR and MISR ........................................................................... 91
Figure 3.8: Average gain over SISR for a varying quantity of utilized frames .............................. 93
Figure 3.9: Plenoptic raw image and magnified images .............................................................................. 95
Figure 3.10: Example of view point images and EPI ................................................................................ 97
Figure 3.11: Raw image captured by the plenoptic camera ................................................................. 98
Figure 3.12: Grayscale raw image and 9x9 VPs ....................................................................................... 99
Figure 3.13: Result from interpolation .................................................................................................... 100
Figure 3.14: VP image cropped and Interplated cropped VP image ....................................................... 100
Figure 3.15: Comparison results for super-resolution Numbers image .............................................. 102
Figure 3.16: Comparison results for super-resolution Toy image ...................................................... 103
Figure 3.17: Comparison results for super-resolution Truck image .................................................... 104
Figure 3.18: Comparison results for super-resolution Flower image ................................................. 105
Figure 4.1: Disparity information extraction ........................................................................................... 112
Figure 4.2: Block diagram of the proposed LFBMSR approach ........................................ 114
Figure 4.3: Directional Adapting Interpolation .................................................................. 115
Figure 4.4: Search window of the LFBMSR algorithm outline ........................................ 119
Figure 4.5: Mapping and 4D transform .............................................................................. 120
Figure 4.6: Light Field Hybrid Super-resolution method .................................................. 122
Figure 4.7: Raw and View point Images (VPS) .................................................................. 126
Figure 4.8: Magnified Numbers image .............................................................................. 128
Figure 4.9: Magnified Toy image ...................................................................................... 129
Figure 4.10: Magnified Lego image ................................................................................... 130
Figure 4.11: The output Numbers images of all competitive methods and close ups .......... 132
Figure 4.12: The output Toy images of all competitive methods and close ups ............... 133
Figure 4.13: The output Lego images of all competitive methods and close ups .............. 134
Figure 4.14: The output Truck images of all competitive methods and close ups ............. 135
Figure 4.15: The output Flower images of all competitive methods and close ups .......... 136
Figure 4.16: Enhanced resolution image obtained using the proposed LFBMSR method .... 137
Figure 5.1: Raw and view point images, decoded (Alam and Guntrunk, 2017) ................. 150
Figure 5.2: Extracting VPs which lead to focused images ................................................ 153
Figure 5.3: Refocusing scheme, method based on VP images .......................................... 155
Figure 5.4: Raw image (Unprocessed) captured with standard plenoptic camera .......... 156
Figure 5.5: From the 9 refocused images extracted these are 1/9 and 9/9 images ............. 156
Figure 5.6: Flowchart of the proposed method ................................................................. 157
Figure 5.7: All-in-focus super-resolved Spiderman images ............................................. 162
Figure 5.8: All-in-focus super-resolved Lego images ..................................................... 163
Figure 5.9: All-in-focus super-resolved Numbers images ................................................ 164
List of Tables

Table 3.1: Average PSNR outcomes HYSR in dB for $N = 5$ ................................................................. 92

Table 3.2: Average PSNR and SSIM for LFHYSR results ................................................................. 106

Table 4.1: Values of Numbers, PSNR dB and SSIM compared among the LFSR approaches 139

Table 4.2: Values of Toy, PSNR dB and SSIM compared among the LFSR approaches ........ 140

Table 4.3: Values of Lego, PSNR dB and SSIM compared among the LFSR approaches ....... 141

Table 4.4: Values of Truck, PSNR dB and SSIM compared among the LFSR approaches ...... 142

Table 4.5: Values of Flower, PSNR dB and SSIM compared among the LFSR approaches .... 143

Table 5.1. Total average of Spiderman, Lego and Numbers PSNR and SSIM results ............ 165
1. Introduction

In order to introduce computer vision and graphics, it has been described that light field (LF) conceptualization is adopted widely in several regions. Such concepts can be deliberated as the core idealization for 3D display techniques, as well as image-based rendering, attributing to comprehensive, simple notions. It an invaluable tool used for building novel applications (for example digital refocusing), computational photography and enhancing the conventional camera’s capability. A renowned example can be taken from the movie, “The Matrix”, which displayed a bullet-time effect by using light-field-centred methodologies. These methodologies have been frequently used in other movie productions, as well as other industries, including the inspections within manufacturing (Skinner et al., 2016; Sun et al., 2016; Chen et al., 2017; Agarwala and Dontcheva, 2004).
1.1 Background

It can be presumed that image super-resolution (SR) of LF cameras has been recognised as a substitute tool used against the commonly applied interpretation of pipeline, drawn by using ray tracing and 3D geometry. A plenoptic camera is a LF camera which has many features compared to the conventional camera, for example the camera can capture images from different angles and the user can select which part of the image to be in focus after the image has been taken. Lastly, the plenoptic camera can extract depth map images. With reference to a captured incidence of light field, a crucial point which has been identified describes multiple light rays participating in the rendering of a single scene by means of diverse perspective reduction for encouraging an interpolating connection between them, as well for the selection of the relevant rays. Furthermore, this phenomenon appears to be attractive under several application settings, as it has been commonly deliberated as a critical matter for digitizing that particular scene in a perfect 3D model, in addition to realistically rendering the attained model photo. Conventional cameras are used for taking pictures and videos where the majority of people care considerably about the quality of their videos and images and less interest in the camera models; typically, only those who specialise in the movie industry care about the features of each camera. For instance, the focus of the movie can be changed after shooting, also the quality of the resolution and depth of field, in order to direct high-quality movies.

1.1.1 Difference between Conventional and Plenoptic Cameras

In its most general sense, there are huge differences between the plenoptic camera and the conventional camera. The plenoptic camera design shown in Figure 1.1, is a LF camera, it
captures images from different view points (VPs), this is due the feature of the new sensor type named microlens array (Sun et al., 2016). These microlens array are located between the main lens and the sensor. Consequently, the output of the plenoptic camera is a raw\(^1\) data that needs processing using simulation programmes such as Matlab.

![Figure 1.1: Differences between conventional and plenoptic cameras. Left: Conventional camera design. Right: Plenoptic camera design (Sun et al., 2016).](image)

First, calibration is applied to allocate the centres of each microlens with the micro images using a white raw image with the LF image. Consequently, from the micro images, sub-aperture images are extracted in order to be generated as an input to the proposed methods.

In Figure 1.2, the multi view images are extracted from the raw image, these are termed “sub-aperture images” or VPs (Adelson and Wang, 1992). However, the images taken by conventional cameras are captured using a single lens, thus the output is a 2D single image. Furthermore, the plenoptic camera records radiance on the sensor plane, whereas conventional cameras are capable of capturing irradiance from varies angles in a space from a single view point.

\(^1\) Raw is an unprocessed data (image) taken by the plenoptic camera.
Figure 1.2: Raw image captured the plenoptic camera at top (Hahne, 2016). 81 VPs extracted from the raw image shown in the middle. Bottom, the central VP image 281 x 186.
In all positions in 3D space, the flow of light is represented by the plenoptic camera as a 4D function of viewpoint (3D), inclining angle of the ray (2D), colour (1D) and time (1D). Furthermore, (Levoy, 2006) has convened that flow geometrical optics is particularly instigated by the rays, however are they are measured as radiance. Moreover, a LF can be described as a function, which is typically concerned with the radiance of the rays while it transports (Michels et al., 2018; Levoy, 2006). It has been observed that the ray has been parameterized by a certain time, position and direction, which encourages the LF to perform in a 4D function. Therefore, each collection of photographs (either one or more) of a particular scene is representing the light field’s sub-sampling, as well as the low dimensional LF segments which can be identified in some images, however, there is the possibility that not all of these can be conventionally identified in the photographs (Georgiev and Lumbsdaine, 2009; Ng et al., 2005; Michels et al., 2018). An example can be taken from Google Street View which can be considered as a large LF, spotting and locating which are fixing a certain point. This conceptual flexibility and comprehensiveness make the LF to appear well suited for the image-based interpretation. Plenoptic cameras are available commercially from some manufacturers, such as Lytro, which are based on the standard plenoptic camera. The camera that has been used for the proposed method is the standard plenoptic camera, also known as the Lytro camera.

1.1.2 Concept of Super-resolution in Image Processing

Generally, a technique is needed in image processing in order to enhance images or videos, this technique is termed as SR. Furthermore, it can be described as the procedure to generate a HR image, using a single image or a number of multiple images that suffer from LR.
Practically, SR is categorised according to the information used as an input, whether it is single or multi frames (Amisha and Suryakant, 2012; Park et al., 2003; Borman et al., 1998; Farsiu et al., 2004; Elad et al., 1999; Nasrollahi et al., 2014). SR resolves the limitations of LR quality images, specifically for military applications that require enhanced videos and images, such as for car number plates; as well as for medical imaging applications.

The next figure, Figure 1.3, presents an example of LR images of multi frame SR, where the LR images are aligned to create HR image. The difference between each LR can be half or one subpixel shift, depending on the camera position. In contrast, the available multi frame LR images data are not applicable in the single image super resolution techniques, because it uses single a LR image which has limited information available to super-resolve one HR image.

The methods of single framed SR are presented in (Dowski and Johnson, 1999; Freeman et al., 2002; Glasner et al., 2009; Kim et al., 2013; Kim et al., 2010; Yang et al., 2010).

**Figure 1.3:** The multi-frame super-resolution concept. The left side depicts the grid that is representing LR images displaying a similar scene containing sub-pixel arrangements. Therefore, the HR image represented in the grid given on the right side can be attained through fusion of the complementary data using SR techniques (Yue et al., 2016).
1.1.3 Light Field Refocusing

Light field refocusing is a feature which the plenoptic camera can provide by selecting any part of the image to be in focus after the image is taken. (Isaksen et al, 2000) demonstrated virtual refocusing from the bright field for the first time. In a recent study of the original LF rendering graphics file, sometimes referred to as a synthetic aperture photograph (Levoy and Hanrahan, 1996; Alexander and Leif, 2009) it is suggested that the interesting change of the inclined focal plane be taken into consideration, for example, angle the film surface toward the main lens. However, these refocus demos have two problems. First, large camera arrays are not suitable for capturing the necessary LF dataset and can be used for long-time scans associated with traditional photography or moving cameras for spontaneous shooting must be used (Wei, 2016). Second, because of the virtual lens aperture, sampling is not complete (namely, due to a camera gap), the results tend to show high aliasing in the blurred region. Subsequently, aliasing is significantly reduced with the optical design by the integration of the light that is passed through the aperture (Eltoukhy and Kavusi, 2003). Additionally, it is an initial attempt for refocus to be formulated exactly based on the actual imaging process performed in the camera as it is closer to traditional pictures than previous works (Ng et al., 2005).

1.2 Limitations of Existing Work and Challenges

The main drawback is that the plenoptic cameras provide limited depth of field (DoF) and deliver images at (significantly) lower resolution than that of classical cameras due to technological constraints. For instance, the final resolution of sub-aperture images extracted from Lytro field camera data is only 300 x 300 pixels (Ng et al., 2005).
In addition, the main lens’ focal length in the microlens viewed in contrast with the sensor to the microlens of the focal lens has been identified to be significantly larger. Consequently, every microlens is capable of receiving a parallel light ray with lower spatial resolution of all the probable guiding rays. Moreover, the integration in all of the pixels of every micro-lens are signifying a single pixel identified in all of the final images observed on a depth plane. The recent approaches applied SR to the plenoptic images and as opposed to applying BM in order to lower the computational complexity, rather it was applied to enhance the depth of the plenoptic images (Farag and Velisavljevic, 2018). Moreover, the recent approaches including (Bishop et al., 2009; Bishop et al., 2012; Wanner et al., 2014) enhanced image details by designing a variational Bayesian framework, by increasing the resolution of the measured LF, which exploits prior knowledge about the scene and estimates both the LF and HR depth map enabling extracts of additional information from available data. Their method provides image formation by distinguishing the plenoptic camera point spread function (PSF) under Gaussian optics assumptions for a several-depth scene. (Wanner et al., 2014) proposed a technique by using the epipolar plane image (EPI) to calculate continuous disparity maps and variational methods to calculate a novel super-resolved image structure of the LF. According to recent studies of this work, none of the above methods exploited depths for image enhancement, however the proposed technique aims to do that. In addition, the proposed method is combining two classical approaches to enhance the final resolution images, as the recent approaches failed to do that. These limitations are solved by aiming to reduce the computational complexity of the LF images, as well as enhance the resolution to achieve better resolution images. First an adaptive interpolation method (Velisavljevic, 2008) is utilized to enhance the image’s resolution of the plenoptic camera by factor 4x4 (Tsai and
Huang, 1984). The enhancement is done through extraction of the plenoptic cameras’ viewpoints\(^2\) (VPs) of the captured raw images. This method is based on directionally adaptive wavelets and can efficiently preserve the sharpness of the interpolated HR images. Second, the proposed methods are using the interpolated images as an input. This work is then extended to apply block matching (BM), which is a technique that is used in video frames to create blocks in several widow in each frame with the intention of computational complexity reduction. The full search algorithm (FSA) approach is considered to be one of the simplest techniques in BM. As it is capable of checking all the candidates separately in the frames and can find a solution with low level error in block matching (Jain and Jain, 1981). In the matter of LF images, it is difficult to reduce the complexity by using a technique such as FSA or other BM techniques because firstly, such techniques are used in video frames and secondly, the neighbouring blocks that sent the information infrequently conduces to miss initiate search areas, creating vectors that are distorted. However, in LF images the VPs haven’t got motion vector as they are extracted from a single raw image. Therefore, this proposed method is taking advantage of the VPs to exploit the depth information and reduce the necessity of larger blocks, while guaranteeing the required robustness. Furthermore, the proposed method uses a block matching approach to extract the sub-aperture image’s disparity information. This approach is borrowed from the recent work of (Sabater et al., 2015) that has been particularly performed well surrounding the depth incoherencies. Such incidence can be advantageous, specifically in explicated image boundaries localization.

\(^2\) In this thesis, viewpoints (VPs) and sub-aperture images are used interchangeably.
1.3 Motivation of Image Super-resolution Applied to LF images

Emerging plenoptic camera technology has attracted significant attention in industry and research, owing to its potential to provide efficient scene depth estimation and to change the focal points without incurring additional computational complexity (Levoy et al., 1996; Adelson and Wang, 1992; Ng et al., 2005; Bishop et al., 2012; Peng et al., 2013; Mitra and Veeraraghavan, 2012). Apply the proposed methods of the LF images allow a better and higher quality of all the LF images. It also allows applications in areas such as the film industry to process images at a higher resolution. The proposed method is a novel concept called LFHYSR which provides a high-quality reconstruction framework; other direct applications also exist, including free VP rendering and refocusing which is shown. Moreover, the method uses free chooses (i.e., open) low resolution (LR) sub-aperture images. The approach consists of two steps: first, the image interpolation technique is applied to enhance each sub-aperture image spatial resolution by 4x4. Second, the HR images extracted using the proposed methods are used to estimate the disparity information using the method of (Sabater et al., 2015). This information is then used to facilitate block matching super-resolution (BMSR) by narrowing down the block search area in neighbouring VPs. Thus, the disparity pattern largely relies on the VPs of the scene from the plane of the LF. Further, the LFHYSR approach, as illustrated in Chapter 3, has been used to obtain the final super-resolved image.

1.4 Research Aims

The aims of this research are as follows: First, develop a novel super-resolution SR approach to enhance the images suffer from LR which captured by LF cameras. Second, SR and interpolation are increased as compared to the LF images. Third, increase the resolution of
multi focused LF images. Finally, create a working SR application in order to allow higher quality to all LF images captured by plenoptic cameras as well as multi images captured by a conventional camera. Moreover, the significant advantage of this application for computer vision is the super-resolved sub-aperture images or the various angles available in a single light field super resolution image which allow depth estimation.

1.5 Contributions

This thesis illustrates the techniques of image SR of LF cameras, from sub-aperture images to multi refocused images (refocusing). It has made a number of novel contributions, which have led to several scientific publications and are explained briefly below.

1.5.1 Hybrid Light Field Super-resolution

A novel light field hybrid super-resolution (LFHYSR) method is used, this combines two classical SR techniques for efficient application to plenoptic images. After this combination, the output hybrid super-resolution (HYSR) image is segmented into the objects of interest. Afterward, sparse representation is applied to super resolve the segmented image. This technique helps to improve the quality by decreasing computations for LF images and extracting significant features from the objects of interest. The gain achieved is demonstrated by the novel method, as compared to the current relevant approaches in terms of both PSNR and SSIM for various enhanced spatial resolutions.

1.5.2 Block Matching Super-resolution for Light field Images

A developed method of the block matching approach is to be used to enhance the super-resolved image of LFHYSR. Lengthening the BM3D filter that will be exploiting the natural local
severances identified in the natural images, is proposed, through considering the 2D redundancies, which are arising in the angular LF dimensions. We will be firstly proposing the creation of the “4D disparity compensated blocks”, that will be comprising of the 2D blocks (highly redundant), also we will be stacking them with similar 4D blocks on a 4th dimension. In addition, the 4D blocks, attained by this process, will be further processed for 4D-transforms domain, here the 4D blocks primary signal will be represented sparsely. Moreover, the 4D transforms will be further combined with the a 2D angular transform, 2D spatial transform, as well as 2D transform which will implied on the 4th dimension that will correspond with the similarities.

1.5.3 Light Field Image Refocusing

Light field cameras configured with the proposed SR approach can be used to record the LF and provide maximum direction resolution. With this simplification it is easy to explain the refocusing algorithm and the performance characteristics. The complexity of the normal LF camera configuration is also considered. The implemented SR approaches that has been proposed in this work are extended to be utilized to enhance the images of LF, specifically the final all-in-focus LF images. The technique proposed takes the advantage of the multi refocused images that are extracted from the raw image and interpolate the refocused images to a chosen target. Thus, apply the LFHYSR form chapter three to increase the final all in focused image.
1.6 Structure of the Thesis

Chapter 1: Presents a summary of light field photography, overview of limitations of recent work, motivation of plenoptic cameras as well as the contributions and organisations of the thesis.

Chapter 2: Shows a brief history of conventional and plenoptic cameras, as well as the existing studies applied to plenoptic images including super-resolution, refocusing and rendering.

Chapter 3: Shows a novel method which applied to enhance the LF images resolution using a combination of two classical approaches and image segmentation approach. The proposed method is called LFHYSR method.

Chapter 4: Presents a novel super-resolution approach for plenoptic images, by using block matching to lower the computational complexity as well as the LFHYSR from chapter three. The proposed method is called LFBMSR technique.

Chapter 5: Explains the refocusing feature of the plenoptic camera. Moreover, apply the proposed approaches to super-resolve the refocused images to gain all-in-focus super-resolution image.

Chapter 6: Presents the discussion and conclusion of the thesis.
2. Literature Review

2.1 Introduction

The most important topic in image analysis is image super-resolution (SR) as it is a well-researched subject focused on pattern recognition, as well as in image processing for zooming and magnifying. The main idea of SR is the technique applied to generate a high-resolution (HR) image by merging the images that suffer from low-resolution (LR). To show the limitations of existing approaches of SR applied for light field (LF) images, a full study has been conducted in this literature review which addresses and compares the recent SR techniques. As a result, novel methods have been developed to solve these limitations.
This chapter presents a short history of plenoptic camera technology, as it is one of the emerging imaging technologies, and SR applications applied to conventional and plenoptic imaging. The research addressed here is a combination of SR and plenoptic camera principles, therefore SR methods and associated techniques have been analysed. The focus is on the recent research work in image SR applied to LF images. This application is of special interest because of the potential extension of applicability of the camera technology.

The organisation of this chapter will be as follows: Section 2.2 illustrates a historical brief overview to the design model of plenoptic and conventional cameras. In Sections 2.3, 2.4, 2.5, 2.6 and 2.7, the chapter shows the recent techniques used in image SR to enhance the resolution of images which taken have been by conventional cameras, also referred to as “classical SR approaches.” Whereas, Section 2.8, 2.9, 2.10 and 2.11 demonstrate a whole study of the SR techniques used for LF images which were captured by a plenoptic camera, also known as “light field super resolution” (LFSR) approaches. The last section 2.12 contains the discussion and conclusion of this chapter.

### 2.2 A Brief History of Camera Technology

Cameras are changing dramatically and significantly in the world of imaging technology and photography, including collotypes, daguerreotypes, film and digital cameras. The history of the camera can be traced back before the introduction of photography. Western scientists took the word ‘camera’ from the Arabic scientist Ibn al Haytham, who named a similar device ‘alqumra’ in the 11th century (Yu and McMillan, 2004). In the following, the principles of the LF and conventional cameras are explained to clarify the features and the differences between them.
2.2.1 Light Field Cameras

The plenoptic camera is a modern type of sensor that captures a full 4D LF on the sensor plane. It was first proposed in the digital camera theories by (Lippmann et al., 1908). In 1992, a well-known research in this field (Adelson and Wang, 1992), experienced a significant development in the plenoptic camera. This development minimised the problems of stereo matching. The plenoptic camera is based on integral photography principles, which is an auto stereoscopic technique that uses 2D (X-Y) array of tiny lenses to capture a 3D scene. It contains a microlens array focused on the pixel of a conventional image sensor and it can capture a series of view point images (VPs) from different angles, which enables good usability in photography and within the film industry (Montilla et al., 2015; Dansereau, 2014; Fiss et al., 2015; Pérez and Lüke, 2015). The VPs are the images extracted from the raw images taken by the plenoptic camera, as shown in figure 2.1.

![Figure 2.1](image)

**Figure 2.1**: (a) Raw image captured by the plenoptic camera. (b) 9 x 9 viewpoints rearranged as a collection of several views. (c) The central view point image.
The camera is based on a sensor focused behind the microlenses by one focal length (Lippmann et al., 1908), the sensors only measure irradiance. Inclining direction is inherently measured by location of the sensor relative to the system of microlenses. Conversely, 2D conventional cameras only record irradiance from different directions in space at a single viewpoint using the two available dimensions (height and width; pixels along the $x$ and $y$-axis) (Engel and Bershad, 1992). Plenoptic camera was introduced in 1992 as a solution to common problems in graphics and computational photography, including the classical pinhole camera, such as rendering images in multi views and multi focusing. As the credit goes back to (Lippmann et al., 1908), who introduced the idea of microlens array. The most recent developments in computational photography have made it possible to manipulate 4D photographs. In this solution, the LF of the scene is captured and integral photographs (LF images) are produced. Figure 2.2 shows a typical setup of the system design of a plenoptic camera. First, the camera projects the view onto a microlens array. Then, the microlens is extracted through the exit pupil onto a sensor of the camera lens. As the sensor of the camera records only irradiance, the angle comes inherently from the position of the sensor; the rays arrive at plane of the image and the position of the objects is sampled by the microlens.

Figure 2.2: The microlens array has an image projected onto it by the plenoptic camera. The directionality of rays is encoded by the sensor from this point (Adelson and Wang, 1992).
(Adelson and Wang, 1992) designed the plenoptic camera to be used in reducing the connection problem in stereo matching. To make this achievable, the correct place of the microlens array is positioned at the focal lens main plane (i.e., the optical system of the front focal point) and the position of the image sensor is placed behind the microlenses. However, their work has precise and specific list of drawbacks, including the LR in the final resolution of the image. Furthermore, the captured image by this camera system exhibits missing information, including calibration. Therefore, Adelson and Wang stated that the system’s main drawback is that the stereo baseline is confined to the aperture dimension of the lens, which lowers the quality of the image and the depth may not be resolved. Figure 2.3 shows the design of plenoptic camera by Adelson and Wang.

![1992 Adelson - Plenoptic camera](image)

**Figure 2.3**: Camera array of tiny pinholes placed on the image plane can be used to analyse the structure of the light affecting each macro pixel (Adelson and Wang, 1992).

Some commercial prototypes based on this principle have been presented (e.g., Lytro Camera in 2006 and Raytrix Camera in 2015). A variation on this layout, named the “plenoptic camera 2.0” or “focused plenoptic camera” was designed so that the camera sensor as well as the microlens array are capable of being shifted backwards (Adelson and Wang, 1992), so that the microlens arrays relay the image onto the camera sensor (Georgiev and Lumsdaine, 2008;
In 2005, a Lytro camera, shown in Figure 2.4, was designed by (Ng and Yau, 2005). The Lytro camera contains a microprocessor, microlens array and LF sensor. The sensor of the Lytro camera is covered by a microlens matrix. These microlenses provide an additional amount of information which can be measured precisely (e.g. bit). Moreover, this additional information is not provided by lenses, but by incoming light rays. Then, the Lytro microprocessor finds what an image should look like when it is in focus (Ng and Yau, 2005). The authors proposed a new approach to solve the limitations of Adelson’s camera design by interpolating the missing calibration information (Ng and Yau, 2005). In addition to that, the image can be processed using specific software, such as Matlab, after it is taken because the image is unprocessed when it is captured. However, the resulting spatial resolution is low compared to classical cameras.

Figure 2.4: Left: The Lytro camera. Right: Raytrix camera. (Ng and Yau, 2005).

A historical overview is given above about conventional and plenoptic cameras. The next section demonstrates some existing SR techniques and algorithms used previously. Some of these methods are evaluated and compared to the proposed approach.
2.2.2 Conventional Camera

The conventional camera is based on mechanical and chemical processes, which exploits the chemical features of photographic film in sequence. The idea of the camera Obscura design, shown in Figure 2.5, goes back to Ibn al Haytham in the 1500s, in that the light can pass through a small hole in a box and it is projected onto a background without the use of a lens (Schwarz, 1989; Gupta and Hartley, 1997; Zghal, 2008; Paolucci, 2006).

![Camera Obscura](Paolucci, 2006)

In 1826, Joseph Niepce developed the camera Obscura and created a permanent photograph, this step simplified the power of the photograph (Paolucci, 2006). Later in 1839, the time taken to take a photograph was reduced from eight hours to a few minutes. Credit can be attributed to Daguerro who invented the ‘Daguerreotype’ camera. One of the main characteristics of this camera was that the image was projected onto a bright mirror with a metallic silver surface, so it could perform as either minus or plus (Horstmeyer, 2009 and Yu and McMillan, 2004). The example shown in Figure 2.6 is the design of the Daguerreotype camera.
By the 1850s, it was common to see the improvements in the design of the camera which lead to create doing a mobile photo studious (Robinson et al., 2017). During the 1870s, the dry plates were released came out. George Eastman invented the first simple Kodak camera, which was factory based and pioneered using photographic film. Eastman’s camera was designed on a basic concept with a single shutter speed and a fixed-focus lens. Kodak later exploited Eastman’s camera in the 1880s (Mess, 1948). In 1924, a new camera design shown in Figure 2.7, called the 35mm was invented by Leica, in Germany, which revolutionised photography by taking clear nice photos, however it has a drawback as it was heavy (Day and McNeil, 1996).
In 1949, Canon manufactured one of the best conventional camera formats’ globally, as well the first single lens reflex (SLR) camera and the first digital image, which has become the industry standard for professional photographers. It allows photographers to precisely see the captured image by a system which includes a mirror and a prism inside it (Hall, 2018). The system has a single light path, however the design of earlier cameras had two, one through the lens to the viewfinder and one to the film through the lens. By the mid-1980s, many companies had been working on the design of digital cameras. Canon is considered the first to present a viable prototype, revealing the digital camera in 1984. However, it is also acknowledged that Canon never manufactured the digital camera and sold it commercially. In 1990, Dycam model 1 was the first digital camera to be sold and was purchased for approximately $600. In the following year, Nikon F3 was the first DSLR whose body was attached, with the storage unit manufactured by the Kodak Company. However, it was not until 2004 when digital cameras had become dominant in the market and were outselling film cameras with strong capture in the markets. The latest camera design by Canon, which is among the three best cameras in 2018, is the EOS 5D MARK III (Hall, 2018), shown in Figure 2.8. The key features of the camera is to allow professional photographers to shoot images continuously in high speed and maintain high performance, even if the light is low. At the most fundamental level, the design of the camera has not changed since the 1850s, all cameras are built in the same fundamental chassis, that is a lightproof box with a hole in the side. As the camera has developed, it is no longer used for scientific research alone and has started to be used for art and documentation also. In a sense it changed the world.
2.3 Existing Classical Super-Resolution Methods

As mentioned above, the SR method can be defined as the process of creating a single HR image using a single LR image or a set of multiple LR images (Jung and Hwang, 2004). The techniques of SR can be split into two classes: First, interpolation methods of nonlinear and linear, which are explained in section 2.4. Second, multi or single frame approaches which are based on SR (Boulanger et al., 2007; Protter et al., 2009; Farsiu et al., 2004; Shi et al., 2014; Karimi et al., 2014). SR techniques can be divided into three sub-classes: 1) “Multi-frame super-resolution (MISR)” (park et al., 2003), 2) “Single image super-resolution (SISR): Example-based super-resolution” (Taniguchi et al., 2014: Yu et al., 2012; Freeman, 2002; Winter et al., 2015; Xian et al., 2013; Yu et al., 2013). 3) Interpolation-based methods (Hu et al., 1992; Kumar and Liyakathunisa, 2010; Li et al., 2010; Wei, 2016; Yang et al., 2010).
2.3.1 Multi Frame Super-resolution

Since significant work proposed by (Tasi and Huang, 1984), SR reconstruction is becoming an emerging area of research. Two decades ago, different techniques had been proposed, representing approaches from signal processing to machine learning perspectives and carried out in the frequency or spatial domains (Freedman et al., 2011). This section explains some of the multi-frame techniques.

There are several techniques to reconstruct an image from the same amount of data by combining multiple LR images in order to obtain a HR reconstruction, such as the MISR technique. However, the classical single image SR techniques use a single image to increase the resolution (Park et al., 2003; Tasi and Huang, 1984) whereas the MISR technique is a higher quality technique compared to SISR, as it allows for an extra amount of data to be selected. Alternatively, some methods exploit motion in image sequences to enhance the resolution. For example, video sequences and frames that have motion can be reconstructed based on these motions, using the multi-frames taken. A different way is to exploit the knowledge about the image by forming models and estimate model parameters given a limited aperture; the image is blurred by a known kernel. In this case, blind deconvolution or motion deblurring are used to restore the image information (Chakrabarti, 2016; Dong et al., 2017; Yan, et al., 2017; Pan et al., 2016). Another possible way is to compress sensing using random sampling (Tasi and Huang, 1984). The LR images used in the MISR technique must be aliased, sub-sampled and aligned to a sub-pixel degree of accuracy (Tasi and Huang, 1984; Boominathan et al., 2014; Pan et al., 2016). Furthermore, this precision is necessary as LR images taken, which vary by integer units, contain identical information and no new pixel data, therefore the SR technique will not be active at producing a SR image to the input images. Thus, the HR images are not
obtainable from the LR images, if their subpixel shifts vary from one another and if aliasing is not done. For this reason, a HR image can be created after collecting the new information from each LR image (Tasi and Huang, 1984).

To get various angles of the same view, motion from the related scene need to occur from frame to frame via video sequences or multi-scenes. A multiple scene can be created through a single camera with different viewpoints or through several cameras positioned in various locations. Controlled motion of the imaging device is an effective means to provide these scene shifts, for example, a camera mounted on a satellite orbiting the planet will have a predictable velocity and this information is capable of obtaining the alignment of multiple images within a single scene. The same principle can be applied to the uncontrolled motion of either the imaging device or objects moving within the scene. If the motion is measurable or could be computed with sub-pixel precision, SR techniques might be applied subsequently.

Another technique of multi-frame super resolution is super-resolution reconstruction (SRR). (Glasner et al., 2009; Yang et al., 2016; Yeganli et al., 2016) state that SRR can be split into two techniques, the traditional example-based SR and MISR. As explained above, MISR is a technique applied to generate HR reconstructed images by using a collection of various images that have LR, as well as the same scene, which is a necessity. Nevertheless, in example-based SR a database is used containing LR and HR image patches. Usually, these patches have a relative factor 2 scale in HR and LR image pair. Moreover, this information is utilized with regard to form a new LR image in an attempt to reconstitute a HR image. Over time, higher SR factors can be obtained by applying this process multiple times. To reconstruct the HR image, the classical multi-image SR algorithms are used. However, in the case of example-based SR, the image will not assuredly be reprocessed precisely (Kumar and Liyakathunisa, 2010).
2.3.2 Example Based Super-Resolution

Example-based SR is a learning-based approach to create a HR image, based on the corresponding relation in a set of training LR and HR image pairs, prepared in advance (Taniguchi et al., 2014; Yu et al., 2012). There are several methods of example-based SR: some of these approaches use sharpening and smoothing, as well as interpolation. The techniques for smoothing are Gaussian, Wiener and Median (Tai et al., 2010). In the previous proposed techniques, a number of frames were used for example-based SR and the image reconstruction contained paired spatial data (Yang et al., 2010). However, in most circumstances, only a single LR image is used, thus the methods for multi-frame images are not effective. Most recently, novel example-based SR approaches are presented to solve the limitations of the quality and imaging measurement; several examples are provided in the section below. Unlike earlier techniques which used a specific rule from the captured image to identify parameters, these techniques estimate the selected image from other images using a local sampling method (Xu et al., 2012).

2.4 Interpolation-based Methods

(Sun et al., 2013) states that image interpolation is a well-known approach for processing video, as well as images. Different techniques have been designed in image interpolation to accomplish detail-preserving image scaling and high-ratio. The basic processing flow can be illustrated and summarized in Figure 2.9. The methods of image interpolation methods can be broken down into different categories:
2.4.1 Non-linear interpolation methods

The methods of non-linear interpolation features depend on the implied detection of the local image. (Li and Michael, 2010) stated that non-linear methods solved the drawbacks related to the linear interpolation. Moreover, new edge directed interpolation (NEDI) is considered a unique effective method among non-linear techniques, which apply regulated geometric co-variance in order to acquire HR image sizing (Pérez et al., 2003). Instead of following the methods proposed by several studies to enhance the efficiency and quality of the NEDI process, these particular methods are still restricted to the $2^n$ ratio, in which $n$ is an integer. A huge computation cost is also incurred in these methods, which makes it hard to construct real-time implementation for the application systems of digital TV.

2.4.2 Linear Interpolation Methods

For the reference image, the linear interpolation methods are also frequently use the convolution kernel. Through calculating a linear combination of surrounding pixels and exact values, the new value of the pixels is able to gain. (Keys et al., 1981) proposed the inclusion of
bicubic interpolation, nearest interpolation and bilinear interpolation methods of linear interpolation (Pérez et al., 2003; Shi et al., 2014). These are considered to be computationally efficient, mainly as the bicubic interpolation presents better visual resolution images. These methods are also able to retain low frequency content in the processed image, which better enables these methods to indulge in detail.

2.4.3 Directional Adaptive Interpolation

The matter of obtaining a HR image through its LR counterpart is addressed by the image interpolation (Jung and Hwang 2004; Yeganli et al., 2016). It is based on various applications of the real world, including police work, security and life science, in which the captured picture’s resolution in aim is low and the arising effects mainly seem to be attributable to a restricted set of device pixels that are couple-charged and utilized in industrial conventional digital cameras. In addition, the pixels are defined as the exact and actual image resolution, whereas the effective resolution is can be considered difficult to define as it relies on the subjective judgment of human opinion and perception. According to the proposed linear and non-linear image interpolation techniques, the list of widely used methods to increase the images’ resolution includes the nearest-neighbour, the linear methods, bilinear and bicubic interpolation methods (Thvenaz et al., 2000). Approaches of image interpolation used to vary on the basis of the algorithms mentioned, in order to magnify the LR images. Both techniques are adapted locally through wavelet-based image interpolation and the directional adaptive interpolation (Antonini et al., 1992). Both of these methods are capable of enhancing the resolution by 2x2 and 4x4 factors. There is also a difference in their performance, for instance the method of directional adaptive interpolation can be implemented for the extraction of
edge information, alongside several indications and directions from the image of LR and the sharpness of HR image. Detail is preserved through the processed image and its utilisation of the information (Velisavljevic and Coquoz, 2008). However, the locally adaptive wavelet-based image interpolation considers making images at various resolutions over a LR single image. The sections below comments on the present, multiple techniques of interpolation in detail.

2.4.4 Locally Adaptive Wavelet-based Image Interpolation

The image interpolation, locally adaptive, wavelet-based approach depends on the directionlets, which is also known as multiple-direction wavelet transform. Based on earlier studies, the LR image is extracted from a HR references image and appears as a “low pass” output of the 3 band 2D wavelet transform (WT). A related approach is utilized by (Mallat and Zhong, 1992). The “low-pass filter” estimates the missing information from the two high-pass sub bands and the HR low-pass from the present LR image. This would enable the use of the inverse 3-band 2D WT in sub bands that could outperform the previous studies and lead to the presentation of a HR reconstructed image, along with preserved sharpness. The process of the proper method implemented is shown in Figure 2.10, which also applied HR through the present LR image (Chang and Cvetkovic, 2006).

Figure 2.10: Block diagram of image interpolation locally adaptive wavelet-based (Chang and Cvetkovic, 2006).
2.4.5 Directional Adaptive Interpolation

The Directional Adaptive Interpolation technique is presented by (Velisavljevic, 2008). The initial step implemented was to divide the image into parts and segments related to the size 16x16 pixels. After this process, each part in the directionlets is then processed on the basis of each set of direction forming a pair.

\[ D = \{(0^\circ, 180^\circ), (0^\circ, 90^\circ), (0^\circ, -90^\circ), (180^\circ, 90^\circ), (180^\circ, -90^\circ)\} \] (2.1)

Utilising “biorthogonal”, “9-7”, “1-D filter-bank” (Antonini et al., 1992). From the surrounding segments the pixels were used in order to filter the borders related to segments, which would avoid the possible effect of blocking in the process of transform caused by the multiple tiny segments. To achieve better direction, pair \( l^*r \in D \) is selected for each indexed segment by \( r \) as:

\[ l^*r = \arg\min \sum_i \left| W_{r,i}^{(l)} \right|^2, \] (2.2)

In this equation the \( n \)th segment is being applied by directionlets, with the combination and pair of \( l \) indicating direction for constructing the wavelet coefficients \( W_{r,i}^{(l)} \). The energy is reduced by the directional map in the high-pass sub bands, which is then determined through a set. It results in the provision of the proper matching among the transform across segments and locally dominant directions. For more simplicity, the direction pairs of the horizontal and vertical are distributed for the segments, which is required to be smooth with the elimination of the deceptive dominant direction. Good quality images among the bicubic, adaptive interpolation and directional adaptive interpolation are displayed in Figure 2.11.
Figure 2.11: Baboon close up images of are magnified using three classical approaches: bicubic, locally adaptive interpolation and directionally adaptive interpolation. a) Shows the reference high resolution image and b) the bicubic image shows blur and noise. However, c) shows a higher resolution than the bicubic and d) is the highest and closest to the reference in regards of visual efficiency (Velisavljevic, 2008).

As a result of studying the recent images interpolation methods, the outcome has helped to determine a powerful interpolation technique, such as directional adaptive interpolation in order to magnify the LF image’s resolution before the proposed novel methods are applied. All details are presented in the coming chapters.

2.5 Point Spread Function (PSF)

This PSF has been mentioned many times in recent LF studies, because it shows an essential role in image formation theory, which helps in the process of creating the whole acquired image. $PSF_I(x)$ consists of two components:
\[ PSF_i(x) = (a_i * w_i) \] (2.3)

\( w_i(x) \) represents the blurring caused from the lens. \( a_i(x) \) samples the spatial integration of the sensor and \((*)\) is the 2D convolution operator. (Delbracio et al., 2012) state that the first-order squared Bessel function is approximated by the diffraction-limited optical transfer function, as well as, typically, the blurring factor. \( w_i(.) \) is divided into an approximated defocus factor by a pill-box function. (Delbracio et al., 2012) declared that the spatial integration function is:

\[
    a_i(x) = \begin{cases} 
        \frac{1}{s^2} & \text{if } |x| \leq \frac{s_i}{2} \text{ and } |y| \leq \frac{s_i}{2} \\
        0 & \text{otherwise}
    \end{cases}
\] (2.4)

The pixels' photosensitive areas are square, where \( S_i \in [0,1] \) is the photosensitive area width. Calculating the camera’s PSF is thus a complicated task that is based on a varied parameter that clarifies the size, shape diffraction and defocus effects of the pixels’ photophobic areas. Practically, it is easier to adopt an easy parametric form for \( PSF_i \), such as, a Gaussian, consequently the parameters are estimated empirically. According to (Chang et al, 2011), the image source of the light point that is targeted far away can be estimated because the sensor’s PSF defines the image of a secluded point target located on a black background consistently. The outcomes of the reconstruction-based method for many enlargement factors can be seen in Figure 2.12, which provides knowledge of translations and PSF.
**Figure 2.12:** The reconstruction-based algorithm. The Gaussian translated multiple times and down sample are used to blur the original HR image. The translation and PSF are the algorithm used. The images shown on the right are compared to the input images resolution. The performance degrades when increasing the magnification, but it seems dramatic (Chang *et al.*, 2011).

### 2.6 Hybrid Super-resolution Methods

Hybrid in image SR is a term which can be used when merging two methods, two datasets or two cameras together. The recent hybrid methods including (Paul *et al.*, 1996; Xian *et al.*, 2015) applied to image SR have used two images as an input, it could be HR with HR or LR with HR. The approaches proposed by (Zheng *et al.*, 2017) (Wang *et al.*, 2018) (Zhao *et al.*, 2018) has recently been presented hybrid approaches but it was using additional hardware with the images. In Figure 2.13, An example of hybrid super resolution (HYSR) is presented.
Figure 2.13: Scheme of the hybrid example-based method, a LR and HR image is reconstructed via external dataset through the regression model. The “Gaussian mixture model (GMM)” is used to guarantee efficient and pointed information and on the other hand, the input feature extraction of LR and HR prior (Xian et al., 2015).

2.6.1 Classical Image Super-Resolution and Observation Model

The process to analyse the reconstruction issues associated with SR are used to verbalize the model for observation, this defines the association with the HR images which are used to observe the images that suffer from LR. Moreover, the observation model has been resolved in a variety of research (Komatsu et al., 1993; Park et al., 2003; Kim et al., 2010) and they provide a wide scope for dividing these models that are linked with video sequences. However, all these approaches are proposed to define the elementary framework associated with the reconstruction techniques of SR, which further defines the model for reconstruction linked with still images and are employed below. Although, this is found to have a direct approach which is proposed to extend the model for still images linked with LF images. The mathematical Eq. 2.5 presented below shows the steps involved to transmit the HR image $x$. 

\[ \text{Feature Extraction} \xrightarrow{\text{GMM}} \text{Regression Models} \]

\[ \xrightarrow{\text{Feature Extraction}} \text{Proxy High-Resolution Image} \]

\[ \xrightarrow{\text{Reconstruction}} \text{Refined Gradients} \xleftarrow{\text{Pyramid Self-Refinement}} \]
desired, linked with the consistent kth which is observed LR frame $I_k$, (Elad and Feuer, 1997; Nguyen et al., 2001) in preparation for datasets. It can be also considered as the process that occurs before the use of the LR image. Then, the SR is the inverse problem that is trying to reconstruct the original HR image which is explained after these equations.

The process to lower the resolution of the HR image is downsampling, image blurring, rotation and warping (Komatsu et al., 1993; Tasi and Huang, 1984). The HR image desired is given as size.

$$X_1 Y_1 \times X_2 Y_2$$

(2.5)

A lexicographical vector written in notation as:

$$\textbf{x} = [x_1, x_2, \ldots, x_N]$$

(2.6)

where

$$N = X_1 Y_1 \times X_2 Y_2$$

(2.7)

where $\textbf{x}$ represents the downsampled image. The directions of vertical and horizontal downsampling factors are the parameters represented by $X_1$ and $X_2$. Therefore, each corresponding kth detected LR image has LR, its size $Y_1 \times Y_2$, as well as the image $I_k$ feasibly denoted as:

$$I_k = [I_{k,1}, I_{k,2}, \ldots, I_{k,M}]$$

(2.8)

for

$$k = 1, 2, \ldots, P$$

(2.9)

and
\[ M = Y_1 \times Y_2 \] (2.10)

On the other side, an assumption has been made in which the variable ‘x’ is found constant during the attainment of various images having LR, regardless of the degradation and motion which is permitted by the model. Moreover, the outcomes observed for LR images which were found with blurring, subsampling and wrapping operators were performed in the images having HR, i.e. ‘x’. However, it has been assumed that each image with LR has been corrupted due to the additional noise. The model was designed for introducing the LR images from HR images, these is represented by Eq. 2.11.

\[ I_k = P B_k W_k x + n_k, \text{ for } 1 \leq k \leq P \] (2.11)

Where \( P(Y_1 Y_2)^2 \times X_1 Y_1 X_2 Y_2 \) represents the subsambling matrix, \( B_k \) represents the blurring matrix \( X_1 Y_1 X_2 Y_2 \times X_1 Y_1 X_2 Y_2 \); \( W_k \) represents the warping matrix of size \( X_1 Y_1 X_2 Y_2 \times X_1 Y_1 X_2 Y_2 \) and \( n_k \) denotes an order noise vector lexicographically.

**Figure 2.14**: Process of downsampling images from HR to LR, is called the observation model (Park et al., 2003).
The process of SR is opposite from the observation model explained above, as typically in order to apply SR, LR images are used as input that are already captured with sub-pixel shifts. After down-sampling is complete, an interpolation approach is used to map the LR images and create an SR image, which is the most intuitive method for SRR. Figure 2.15 shows the three stages applied after LR is observed, in case of multi images registration is applied. However, in a single image there is no information to register from different images. First, the image registration process is applied, which means pixels from the LR image are all mapped back into the reference image. Then, interpolation is used to create a higher quality image resolution. Finally, the image restoration and deblurring is applied to the reconstructed image to clear the affected blur and noise of sensor PSF.

![Figure 2.15: Scheme for super-resolution (Alam et al., 2009).](image)

2.6.2 Approaches of the Single Image Super-Resolution

Enhancement of the images from LR to HR is the process of SISR. The method used in (Kim et al., 2010), is modified by (Bätz et al., 2015). Their method is to estimate high-frequency details by using a kernel ridge regression (KRR). In Eq. 2.12, the KRR is adopted to learn the map among the LR input and the targeted image. Furthermore, the ideas of gradient descent and
kernel matching pursuit (KMP) are combined with the intention of finding the regression scheme sparse solution and to decrease computational complexity. Through utilising a natural image prior model for the purpose of reducing the resounding articles established through regularization, the transitional outcomes are additionally post-processed. The image prior technique is a part of a conventional neural network applied to increase the resolution of an image itself with no prior training. To reduce the cost functional, the authors (Kim et al., 2010) calculated the distance between the output of the KRR localised and the function output

\[ C(x_j) = \| K^T_{bx(1,n_j)} A_{bx(1,n_j)} g_j(x_j) \|^2, \quad \text{for } j = 1, \ldots, l, \]  

(2.12)

where \( C(x_j) \) represent the output function of the KRR and \( g_j = g_j(x_j) \) is the centre input \( x_j \) of the localised KRR. This is acquired through gathering the nearest neighbours (NNs) which are used to obtain \( x_j \), in addition to training the complete KRR grounded solitarily upon the NNs. Afterwards, the points of candidates are selected serving as the data points for training conforming to the \( l_c \)- biggest \( C \) values. This SISR approach is based on the SR image interpolation, which is inserted within the digital models of a picture. This has long been studied by researchers, though only lately with sampling approaches and machine learning. For instance, bicubic interpolation is a popular approach to image interpolation; however, it suffers from blurring edges as well as jagged artifacts. The attempts used recently to enhance bicubic interpolation have gain limited achievement. The researchers (Schultz and Stevenson, 2007) used a Bayesian approach for SR; however, it hypothesizes instead of learns the prior probability. Subsequently, (Bätz et al., 2015) implemented a hybrid SR imaging system and used the SISR technique to merge it with MISR in order to get HYSR.
2.6.3 3D Hyperspectral Images Super-resolution using Sparse Representation of the Bayesian

Notwithstanding the established effectiveness of hyperspectral imaging in several tasks of computer vision, its increasing applicability and use are hampered by the fact that it has a low spatial resolution, which comes as a result of limitations in hardware. Research conducted in (Akhtar et al., 2015; Winter et al., 2015) proposes an SR approach on hyperspectral images, which entails fusing images with a HR and hyperspectral images of LR through the use of Bayesian sparse representation that is non-paramedic. The approach, in this case, works by inferring the probability distributions for the spectra of materials in their proportions and scene. After that, the distributions are applied in calculating the sparse codes for the project’s image of HR. Thus, (Akhtar et al., 2015) proposes a strategy of sparse coding that is generic. According to the theoretical analysis (Fotiadou et al., 2016), the strategy intended to find out if this code was accurate enough. The codes that have been obtained are applied by the scene vision projected to create the hyperspectral image with SR features.

The vision features of hyperspectral imaging have recently been purported to enhance the performance of several projects in computer vision, including recognition, tracing, as well as classification, document analysis and segmentation. (Akhtar et al., 2015) affirm that the features have played an integral role in the sector of imaging in medicine related fields, as well as remote sensing. Hyperspectral imaging obtains a realistic representation of spectra on the scene through assimilating its radiance against many basic functions that have been well organised, in terms of spectra. Nevertheless, the current hyperspectral systems are not used or available in the spatial resolution. This is one of the major reasons why they are not employed widely in many projects. Thus, a simple mechanism of utilising sensors with HR is
not applicable, as it diminishes the density of the different photons that reach the sensors, which have already been restricted by the instruments with a high spectral resolution.

As a result of being limited by hardware, the approaches based on software for the hyperspectral SR of images are taken to be very attractive. (Dai et al., 2017) showed that the spatial resolution of the systems using the images by grossly quantizing the radiance of the scene is way higher than the hyperspectral ones. Thus, (Kwa et al., 2017) recommends fusing the spatial information from the images, that the systems obtained, with the hyperspectral ones in the same scenes, through the use of non-parametric Bayesian sparse representation.

The above approach works by fusing the HR hyperspectral images in a process that entails four stages. In the first stage, there is an inference to the distribution probabilities for the reflectance spectra of materials used in the provided scenes and a group of Bernoulli distributions, showing their sizes in the resulting image. After that, it gets to estimate a dictionary and then do a transformation of the dictionary, as per vision quantification of the image with a HR. In the subsequent third phase, the distributions and dictionary are applied in sparse code’s calculation of the image with HR. For that reason, (Akhtar et al., 2015) proposed a strategy for Bayesian sparse coding to be applied along with the Bayesian dictionaries obtained with the Beta process. However, a theoretical analysis of the strategy proposed was conducted to determine its accuracy (Dai et al., 2017). Lastly, the calculated codes were applied with the presumed dictionary to create the hyperspectral image with SR.

The suggested approach helped to improve the results obtained and maintain the advantages of the non-parametric Bayesian Framework, more than the normal approaches that are based on optimization.
For 20 years, the use of hyperspectral sensors has been a common practice in the process of imaging. However, with constraints like budgets and other technicalities, it remains a great challenge to attain hyperspectral images with a HR. This has been proven after extensive and thorough research in the field, more so in the aspects of remote sensing. To ensure that there is a better spatial resolution, fusing of the hyperspectral images with images which have HR is common. Thus, in this case, there is the use of conservative methods which are mainly based on substitution and projection, inclusive of the principal component analysis and intensity hue saturation.

Consequently, the authors (Dai et al., 2017; Akhtar et al., 2015) have used human vision sensitivity to fuse and luminance of the images with a HR, along with the hyperspectral images. Nevertheless, the technique can also lead to defects in the image that is created in the end.

The authors (Kwa et al., 2017; Fotiadou et al., 2016) have accomplished the separation of enhancements used in spatial resolution for images with a HR. Nevertheless, the techniques they employ dictate that the image spectral resolutions to be fused are in good proximity. At the same time, the techniques are not effective in scenarios that are highly mixed. The sliding window technique has been applied to improve their performance.

The latest work done by (Fotiadou et al., 2016) has aimed at investigating matrix factorisation using SR of hyperspectral images. The methods created under the framework work well by fusing the RGB images with a HR to the hyperspectral images. However, the approach in this regard needs experience and enough information on the s transform amongst the images under fusing. The use of methods based on matrix factorisation has clearly indicated very high-quality results of SR in different hyperspectral images. Nevertheless, as indicated by (Fotiadou
et al., 2016), they have a sensitive performance, with the sensitivity geared towards algorithm parameters, more so the matrix size where the factoring of images is done. At the same time, there is no standard way that can be used in incorporating knowledge to improve the methods’ performances.

2.6.4 Hybrid Spatial-Spectral Super-Resolution Representation for X-Ray

The concept is growing in popularity and use because of its non-destructive capabilities. In the use of high-quality spectra of X-ray Fluorescence, it is important to gain enough information on the minor and major essentials used for characterisation and analysis, in order to prove some vital facts. Nevertheless, there exists a relationship between signal-to-noise (SNR) and the X-Ray Fluorescence (XRF) Scanner’s spatial resolution of the spectrum in every pixel, because the time for scanning is very limited and never sufficient (Klepikova et al., 2016; Deng et al., 2017).

For some time now, laboratory systems that use XRF have changed into portable and lightweight items, all because of technological improvements in both the X-ray detection and generation. Essential information which resolved spatially is offered through scanning the sample’s surface with a collimated or focused beam of X-Ray, with dimensions in millimeters, and then analysing the produced radiation of fluorescence in a way that is not destructive, popularly known as macro X-Ray Fluorescence (MA-XRF). The latest versions of the XRF spectrometers are applied in the cultural heritage field to research more on the technology of manufacturing and identifying its source and originality for the artworks Deng et al., 2017).

Just like the other methods of imaging, the use of high spatial resolution, as well as SNR, is suitable for any XRF systems for scanning. Nevertheless, the time of acquisition is normally less, leading to complications between the spatial resolution, dwell time, as well as the
expected quality of the image. For the part of scanning mappings in large scale, a decision can be taken to bring down the dwell time and then enhance the step size incrementally, leading to XRF images with low spatial resolution, as well as low XRF spectra.

Figure 2.16: (a) The XRF map clarifies the Cr Ka distribution on the side of the bedroom. (b) 10 maps automatic registration layers added to RGB’s original resolution (Deng et al., 2017).

A demonstration of the XRF scan is as Figure 2.16 (a) shows. The image has been coded in colour to make it better in terms of visibility. The image was created from 4096 channels and were attained simultaneously through the use of the Bruker M6 instrument used in scanning. One of the image’s main feature is a LR (which measures 96 by 85 pixels). The time taken to acquire it was approximately one to two hours. Considering the fact that the given painting measures 73.6 cm by 92.3 cm, it required a total of 10 such patches to ensure that the entire painting had been captured fully. (Deng et al., 2017) reaffirms that it would also be good to use or obtain a higher image resolution to show the general public and other interested stakeholders. This would make the process of acquisition unfeasible and thus, prevent the use of XRF apparatus for scanning as imaging devices for the HR widefields. Consequently, Figure
2.16 (b) above demonstrates a total of 10 XRF maps registered automatically and then placed on top of the unique RGB picture (Deng et al., 2017).

(Deng et al., 2017) suggest an approach for SR to attain XRF images with a HR by means of the use of a traditional HR RGB image. Their proposed image SR algorithm of XRF images is capable of being applied in spectral images acquired by different scanning techniques, such as the scanning electron microscope (SEM) and others. They then design every spectrum’s model, including the XRF spectrum since the unseen section of the painting cannot be visible to the original RGB image, with the capability of being featured in the XRF image.

2.7 Image Registration in Super-Resolution

This section shows the process of registration of images before SR is applied. (Irani et al., 1991; Zitova and Flusser, 2003; Alam et al., 2009; Yan et al., 2017) define image registration as a process of the alignment of multiple LR images containing the same scene, which are captured at slightly various times, from different angles. Image registration plays a significant role in image SR, through configuring most of the same scenes from various angles that are represented by multiple LR images; it is responsible for transforming data into a coordinate system, as well as mapping the corresponding points to the actual points in the original scene (Siu and Lau, 2005). As stated earlier, image registration is often utilised within computer vision, remote sensing and medical imaging. The authors (Zitova and Flusser, 2003) stated that four groups can be found, based on the manner of image acquisition, by dividing the applications of image registration. The first group is several angles, known as “multi-view analysis”. Various angles have the same image scene exploited to gain a better resolution 2D or 3D perspective of the scene image. The second group, several times, referred to as “multi-
temporal analysis” shows various times for taking the same images. The third group is based on various sensors and is described as “multimodal analysis”, which means that the same scene is found with a variety of sensors (i.e., different cameras) (Qureshi et al., 2010). The scenes are merged together to gain a more detailed and higher resolution scene image. The fourth group is the scene of the model images of the scene which are registered. Once the scene has been taken from different viewpoints, times and sensors, the scenes need to be merged together. Registration of images needs different types of transformations, such as planar homographic transformations, biquadratic transformations and affine transformations (Qureshi et al., 2010). Single and multiple shot images that are observed could be captured with various sensors or frames from a video sequence (Boominathan et al., 2013). These images then need to be assigned to a joint reference frame. The procedure of image registration can then be applied to a place of interest in the aligned combined image. The formulation and registration of an appropriate forward image model are the keys to successful SR and consists of accurate alignment (Zitova and Flusser, 2003; Gaidhani et al., 2011).

The idea of the image registration process from multiple LR frames is presented in Figure 2.17. The necessary information among the LR frames is provided by the subpixel motion that makes SRR achievable.

**Figure 2.17**: Image Registration before SR (Tian, 2011).
The below figure of 2.18 presents the diagram of the schematic and steps of the registration process of the image.

1) Feature detection: closed-boundary regions, contours, edges, corners, et cetera, including other closed boundary regions, in the sensed images and ground truth are detected manually or automatically.

2) Feature matching: sensed and ground truth image detected feature in order to establish correspondence.

3) Estimation of model transformation: the estimation of the parameters and segments of “mapping functions aligning the sensed image” by ground truth image (Zitova abd Flusser, 2003).

4) “Image resampling and transformation” (Zitova abd Flusser, 2003): The means of the mapping images alerts the sense image.

Figure 2.18: process of image registration (Zitova and Flusser, 2003).
2.7.1 Light Field Image Registration and Block Matching

The registration process is required to be accurate during the multiple LR images combination process (Yang et al., 2016). Learning classical image registration approaches can lead to an understanding of how BM works in plenoptic images, as this type of data has some constraints, which are useful in reducing the complexity of classical registration. The only way is to exploit disparity information, by generating a depth map for all views to synthesis the view for each desired position of the plenoptic camera data set. The papers in (Wang et al., 2016) and (Seibel et al., 2015) present an instance of one such process. Moreover, a multiple of \( n \) LR images \( I_1, I_2, \ldots, I_n \) are given to find the main points (points of interest), as well as all images that are descriptors to be matched in \( I_1 \) to the main points of \( I_k \) on behalf of every \( k \in [2, n] \).

The perspective transformation \( M_k \) is estimated using the best similar matches between \( I_1 \) and each \( I_k \). Therefore, \( I_k \) with respect to \( I_1 \) is written as follows:

\[
\forall k \in [2, n]: I_k = M_k I_1
\]  

To find the mentioned main points and descriptors, four feature detectors are subsequently presented, which are considered the most used feature techniques in image registration (Liu, 2016):

1) Scale-invariant feature transform (SIFT): Finding main focuses by image region (also known as main point localization) and scale-space-peak determination. Also, main point descriptor. For example, picture gradients in nearby neighborhood of key point. 4x4 cluster of histograms, each with 8 positioning containers (Liu C, 2016).
2) “Speeded Up Robust Features” (SURF): Is a licensed nearby element finder and descriptor. It can be utilized for undertakings, for example, registration image, classification, recognizing object or 3D recreation. It is slightly related by the scale-invariant feature transformation (SIFT) descriptor. The standard adaptation of SURF is marginally faster than SIFT and proclaimed by its inventors to be more powerful against various picture changes than SIFT (Bastanlar, 2010).

3) The points will be matched using k-NN: The measurement of a case is stored singularly, as well as in its groups, based on a measured simulation in new cases (for example, separate capacities). The estimation utilises the KNN as a non-parametric technique, as well as illustrated response until the beginning of the 1970's (Jabbar, 2013).

4) “Oriented FAST and Rotated BRIEF” (ORB): An allegedly, quick and easy way to SIFT.

These approaches are popular in BM methods of computer vision. However, the drawback of these methods is in the implementation process, as it is still complex with a large computation overhead. Furthermore, these search matching methods are used for classical image registration. The key part of this chapter is to build an algorithm to improve robustness and lower the computational complexity. Furthermore, the above feature detector methods were applied to enhance the depth map and were used for classical images that were captured by conventional cameras. However, the proposed method uses images captured by plenoptic cameras as well, as it aims to use BM to search between the best similar match in all viewpoints (VPs) by exploiting the disparities information.
2.7.2 Block Matching High-Resolution Approaches

As mentioned above, the most classical BM approaches have used BM techniques to extract depth information, several examples are presented below. BM without depth refinement is shown in section 2.7.3, as it is an example of the BM process which can be applied to images and videos that don’t have depth of field (DoF).

2.7.3 Block Matching without Depth Refinement

The estimation process of the motion applied using the classical BM approach is shown in Figure 2.19. A series of $I_t$ images is divided into $N \times N$ pixels of non-overlapping blocks. Inside a window to search ($S$) of size

$$(2L + 1) \times (2L + 1)$$

The current frame $I_t$, every block shows the best similar block matched the prior frame blocks. $I_{t-1}$ is resolved, where $L$ denotes the maximum permitted displacement. Motion Vector (MV) is the location disparity between the best matched block within two blocks; the first one is in the prior frame, while the second block is located in the current frame. The perspective assumed in (Cuevas et al., 2013: Cuevas et al., 2015) that BM is an optimization problem intending search for the purpose of finding the best MV.

![Figure 2.19: Capital BM process (Cuevas et al., 2015).](image)
The well-known “sum of absolute differences” (SAD) approach is a benchmark for BM methods (Cuevas et al., 2013). In Eq. 2.15, it considers the location \((a + k, b + h)\) frame within two blocks, the first one is in the prior frame \(I_{t-1}\), while the second block is located in the current frame \((a, b)\).

\[
SAD(k, h) = \frac{1}{MN} \sum_{a=1}^{M} \sum_{b=1}^{N} |f(a, b) - g(a + k, b + h)|
\] (2.15)

The \((k, h)\) denote the displacement vectors, also \(f(a, b)\) represents the image registered grey values, and \(g(a, b)\) denote grey values of the reference image \(M \times N\) for the selected block.

### 2.8 Light Field Super-Resolution Approaches

As it is essential to increase the performance of LF images due to the limited spatial resolution, there have been numerous recent attempts to solve this limitation. However, there is still a huge gap in their methods. For example, (Wanner et al., 2012; Dansereau, 2014) applied a novel technique which only enhances the resolution by factor 4x4 of the sub-aperture images; it can be acknowledged as a valid attempt, but further improvement is necessary to 8x8 and 16x16 factors, when compared to other studies, such as (Dansereau, 2014; Wanner and Golduecke, 2014). A full study of the recent methods applied to the LF image SR illustrated in the following sections.
2.8.1 Applications of Image Super-Resolution to Plenoptic Camera Images

Using the plenoptic camera to capture image data enables better processing potential, as it allows many features, such as refocusing, depth estimation and capture angular, as well as spatial resolution, independently. Furthermore, it eliminates the limitations faced by photographers using regular digital cameras. Unfortunately, classical plenoptic cameras provide limited depth-of-field (DOF), a low spatial resolution and capture images at very LR (Wanner et al., 2012). For instance, the final resolution of sub-aperture images extracted from Lytro camera data is 300x300 pixels. An example the VPs extraction process is presented in Figure 2.20. Due to the way Lytro sampled the 4D radiance of the scene, the plenoptic cameras have a LR in both angular and spatial information (Ng and Yau, 2005). Simultaneously, to sample the angular distribution of the microlenses, manipulation of each microlens under the pixels is conducted (Ng and Yau, 2005). Consequently, (Wanner et al., 2010; Jachalsky et al., 2010) used a computing continuous disparity map and variational approaches to calculate SR novel methods, by using an epipolar plane image (EPI). Subsequently, (Mitra et al., 2012) performed Bayesian inference to achieve LF super-resolution by using a Gaussian mixture model (GMM) for LF patches. The majority of these techniques increased the resolution by a factor of four. (Veeraraghavan et al., 2007) examined the study related to the new design for the heterodyne camera, in which the sensor plane consists of the LF applied with the help of using mitigation. The authors also mentioned that the system advantage is the HR images reconstruction in focused plane. Along with the tested LF, there was also concern regarding the drawback which refers to the use of masks to avoid the light which could reach the sensor and as a result could decrease the signal-to-noise ratio (SNR).
Figure 2.20: 9 micro images, where 8 pixels from every group of 9 pixels are not used (depicted in white). Bottom: Super-resolved rendering into the final sub-aperture image.

Consequently, extracting the VPs views and creating a HR LF image has also be referred to as light field rendering (LFR) in some research (Ng et al., 2005; Georgiev and Lumbsdaine, 2008; Bishop et al., 2009; Wang et al., 2018) This is quite basic in the creation of new views to form a group of images that have been previously obtained, (Keita et al., 2003) used cameras arranged closely to acquire the images. In many practical situations, the proximity and number of the arranged cameras is not sufficient to fuse the right views. The condition of undersampling leads to issues on the image focus in the fused views. The LFR algorithm is quite simple in how it works, in that every value of pixel added on the images gets parameterized
as a data of light-ray and deposited in a database of light-rays. The viewpoint can be fused from the database through selecting the date of the identified light-rays that go through the point of view. In that entire process, the data of light-rays are inserted on the notion that the focal plane is the location of the objects in the scene.

According to (Keita et al., 2003), a major method to deal with the issue, is to apply the information of geometric on top of the database of LF in the given scene. In the method, the scene’s structure is estimated through a focal surface, which entails a group of depth layers or a mesh model in 3D that represents the given structure of the scene in a way that is more precise than just the usual focal plane.

There are numerous researches from when LF was first adopted with the image-based rendering aim. According to the discussion, (Gortler et al., 1996; Levoy and Hanrahan., 1996) highlighted several related problems, including methods of pre-filtering for anti-aliasing and LF rendering, in addition to interpolation and re-sampling for novel view synthesis. However, (Gortler et al., 1996) mentions the proper usage of depth in order to get high quality rendering. The researchers (Buehler et al., 2001; Isaksen et al., 2000) have examined the synthetic aperture refocusing at the front parallel planes through the parameterization of the LF. (Vaish et al., 2005) further studied the extension of refocusing towards the sloping planes. Moreover, (Shum and Kang, 2000) initially studied image-based rendering techniques in a wide aspect on the basis of geometric information being required. The initial techniques are based on a regularly sampled LF, namely those mainly taken from plenoptic cameras on the usual grid. (Buehler et al., 2001) examined the method which is able to contain the unstructured LF, such as input. In particular, are those connected to a specific framework to conceptually distant rendering algorithms, which are known as the rendering of the sampled
LF, by several assumptions of geometrics and the texture mapping of the view-dependent
(Debevec et al., 1996). This, however, tends to determine the generally accurate models of
gold with the number of images smaller in size. In (Davis et al., 2012) the authors
examine the research work related to the interactive system which is able to acquire and
render the unstructured light fields.

The presence of the improper scene geometry can be implemented in order to acquire better
faithful rendering. Moreover, (Isaksen et al., 2000) states that the sparse angular sampling
could be compensated by the approximate-depth proxy process, yet (Gortler et al., 1996)
proceeded further and provide an initial assumption. Whereas, the research of using a rough
depth map is examined by (Wanner et al., 2014) for rendering a LF from a plenoptic camera.
However, the information of depth towards the super resolve LF is implemented by (Bishop
et al., 2009). Rendering through a wide area of viewing positions of the 3D geometry proxies
can be explained through other reconstructed methods. (Zitnick et al., 2004) the researchers
calculate depth maps per view, for the purpose of layered scene presentation, and
implemented the border matting while warping layer. (Chaurasia et al., 2013; Hornung and
Kobbelt, 2009) presented the particle of GPU-acceleration rendering pipeline, that is utilised
for “per-view dense geometry proxies,” by the rendering system in order to provide the
outcome of output views at the rate of interaction from the novel positions. Although,
(Chaurasia et al., 2013) proposed the rendering algorithm based on the image which uses
super pixels, for instance an over-segmented image and its primitive rendering allows it to
manage unreliable depth information. Object boundaries and depth discontinuities tend to be
tracked by these super pixels. The synthesis values are provided to those who are not in
proper reconstructed regions.
According to this method, the novel images are able to synthesize through a light field. Out of the diverse rays these output images are formed in the LF and possess the rays coming from the numerous viewpoints. This is regarded as a key to present a unique level of adaptability that coordinates for the completion of composite output stereoscopic limitations. Multi-perspective imaging was also used by the artists and painters in the history of art as a stylistic tool. However, the related methods have later been examined by the animators in the production of films, for instance for making and drawing background for the purpose of “2D animation” (Thomas and Johnston, 1995). Furthermore, the geometry, as well as the implementation of the multi-perspective imaging was further studied via computer graphics and computer vision; a similar study is explained by (Yu et al., 2010; Wood et al., 1997). Several kinds of multi-arrays were presented in recent years, including general linear cameras (Yu and McMillan, 2004), the push boom camera (Gupta and Hartley, 1997), multiple centre of projection images (Rademacher and Bishop, 1998), and cross-slit cameras (Padjdla, 2002; Zomet et al., 2003). There is no particular related model of camera assumed in this work, but due to the prescribed stereoscopic disparity constraints, the perspectives of images are optimized.

The above approaches are related to the thoughts that proceed them to a more context of utilizing LF, that is able to generate globally optimal output view on the basis of image sharpness. In comparison, (Bishop et al., 2009) constructed an algorithm through increasing measured LF resolution instead of hardware which can exploit the previous knowledge regarding the scene. The methodology related to the restoration of HR images was also presented from them which cause LF data to take from the plenoptic camera. Such sort of systems uses to consist of limitations including the performance will is resulted from the
resolution of sensor of camera as well as microlens arrays. According to sampling theorem, it illustrates the angular and spatial resolution of the recovered light has a trade-off between them. Perhaps, along with the addition the resolution of the image of systems is consider to be restricted on the basis of the size and measures of microlens. The contribution in (Bishop et al., 2009) implemented the details through the designing SR algorithms instead of just increasing the size of pixel, this resulted in the extraction of the additional information from available data. Their influence of contribution presented image formation model from distinction of the PSF plenoptic camera which is under the optics assumption for the purpose of getting more depth various scene. Generally, on the basis of Bayesian approach they presented an SR method according to which they estimated the the image formation model by the Lambertian textural priors as well as predicted HR disparity map.

However, the proposed contributions of this work are as follows. A hybrid method is created by combining a MISR and HYSISR methods, the resolution of sub-aperture images is enhanced by SR reconstruction under Gaussian optics. Second, the block matching-based SR method that can largely reduce the artefacts and improve the precision of shift estimation is applied. In particular, the quality of the image and the disparity map is increased simultaneously as well as the newly developed methods in experimentation is verified.

2.8.2 Applications of 2D and 3D Block Matching Super-resolution

Block matching (BM) is a technique to of creating number of blocks in image frames from video sequence (Cuevas et al., 2013). Recently, BM approaches have been studied in some research such as (Alain and Smolic, 2017; Danieyan et al., 2008; Nasrollahi et al., 2014; Licheng et al., 2016; Marenzi et al., 2017). The authors applied BM approach for LF images as they lie
to VPs freely. Their studies have specifically shown the usage of BM techniques in 3D images. Exploiting BM for denoising the EPIs were discussed in (Dansereau et al., 2013; Baker et al., 2013; Moghaddam et al., 2016) which initially denoises the angular and spatial images. For example, \( (y, t) \) plane represent the angular distribution and \( (x, s) \) denotes the spatial distribution. Then, this initial measure utilizing the corresponding EPIs \( (y, t) \) plane as it denotes the angular distribution. Despite the fact, such approaches only take into account the 2D features of the 4D LF. Nonetheless, the approaches illustrated exclusive performance, however the 4D structure of the LF neglects entirely to bypass the technique of BM3D (Lebrun, 2012) as well as video block matching 4D (VBM4D) (Maggioni et al., 2012; Dabov et al., 2007).

Another similar technique to the proposed light field block matching super-resolution (LFBMSR) approach, is suggested recently (Zhao et al., 2015). The authors use the wiener filter domain (WSD) regularizer, which operates majorly by a scheme of collaborative filtering, BM3D which explores natural images’ self-similarity. This is illustrated in Figure 2.21.

![Figure 2.21: Block Diagram, WSD (Gustavsson et al., 2018).](image-url)
The operation of WSD follows two sequential steps. Both these two steps filter groups that have the same patches from the measurements given by the Euclidean distance. Every step’s outcome is developed by exploiting the patches that have been filtered in their initial spaces and then conducting simple averages for those pixels that possess more than a single estimate. The two steps use filters of different designs on the patches. The first step that gives a basic estimation which is utilized by the second step utilizes a hard thresholding within a 3D transformation domain. On the other hand, the second step that gives the final outcome applies the outcome of the first step to make an estimate using the Wiener filter in the 1D transformation domain. The filter gets applied in the original data input again. This approach is different from (Danielyan et al., 2008) because of the Wiener filter that is used in the second step. The approach enables sharper outcomes to be obtained with clearer details and lowers the cost of computation. The procedure that is employed also includes design elements that enhanced the performance of the system and lowered the complexity of the computation.

Also, WSD is employed in WSD-SR. This needs the filtering strength to be modulated in a manner that it can decrease successfully as it is nearing the steady state. Iterative parameter procedures that are input dependent.

1) A size of a search window that is adaptive

2) Grouping use of information stateful operation

3) Wiener filter within the similarity domain

The decisions regarding the design and the selection of parameters are both discussed in this part. There is empirical evidence for every decision made and presented according to the reconstruction quality and the computational complexity. A scale factor of four was used with tests done on (Zhao et al., 2015). A sampling operator $H$ was also applied to set an antialiasing
filter within the bicubic interpolation. The characteristic under analysis is the one that shifts between columns in the table and the column with the * mark shows the final design. In contract, the proposed BM in this work is an extension of LFHYSR which applies different technique compared to the above-mentioned methods as well as the proposed approach demonstrates that by considering the whole 4D LF structure, it can fundamentally outperform the competing methods.

Another application that uses block matching is multi-colour imaging in neuron tracing. According to (Gustavsson et al., 2018), the secret to understanding the brain function is the connectivity which exists among neurons. Mapping the connectivity between these neurons in brain circuits needs techniques of imaging with a spatial resolution that is high to enhance the tracing of neurons and the molecular specificity for marking the population of molecules. In this case, a three-dimensional SR multicolour imaging method was tested for tracing the connectivity of neurons using hippocampal neurons that were obtained from neonatal wild tripe rat embryos as a system of modelling. Using an approach of labelling that is membrane-specific which enhances the density of labelling as compared to the labeling of the cytoplasmic type, the neural process was captured at 3D 116 nm resolution and 2D 44 nm resolution as it was determined with the consideration of the fluorescent probe’s localization precision and the label density Nyquist criterion. When comparing with confocal images, the results revealed that with the achieved resolution, there was the possibility of distinguishing the neural process in those SR images. The accuracy of the tracing was enhanced with the help of the SR imaging of multicolour. The resolution that has been acquired from this process was influenced by the density of labelling and not the fluorescent probes’ localization precision.
It has been stated in (Cruz et al., 2018) that it requires that there are a 3D tracing process and long distances for neural tracing than those which are provided by the field of view of some microscopes. For this reason, a procedure for 3D imaging was implemented to duplicate the neurons’ process in the vitro and in those areas equal to several perspectives. In this particular case, a sample measuring ~20,000 µm² areas was imaged. There was neutrality in cultures and thus all the procedures were taken by the use of the imaging, a depth of ~1.4 µm that is equivalent to a volume of ~28,000 µm³. The approach used here allows the recreated volume to take any size and can only be affected or limited by the frequency of performing the imaging. With the labelling of high density and the activation powers that were low, it was easy to decide to image only a section of the STORM enquiries for every round of imaging. This, therefore, made it easy to compare the structures found in the overlapping regions. The alignment error was further found to lower as compared to that of the image’s final resolution (Lakadamyali et al., 2012).

2.9 Refocusing of the Light Field Images

The applications regarding the refocusing of LF are described in this section. After a single exposure, the focus of the output image is changed. Extracted properly after calculation, from the LF as seen in Figure 2.22, (a) to (e), it contains 5 images. The images were captured by exposing the prototype LF camera once to exposure of 1/125 seconds (Ng et al., 2005). The images (a) to (e) are refocused of the LF, shows girls with white hat similar to images seen in the camera's viewfinder. The reason behind this view is that the digital refocuses the physical
simulation based on images from real cameras. The focus is on the depth of the output image and a camera that simulates the light flow in this virtual camera.

Figure 2.22 (f) is another image showing the digital extended DoF. Subsequently, to reduce the size of the aperture of lens, conventional cameras can be used that makes it possible to make use of imaging methods that are conventional optically extended DoF. By gathering the sharpest parts of the images, image (b) can be calculated for (a) to (e) and it can also be considered as the refocusing of each pixel, closest to the considered direction at the object’s depth.

![Figure 2.22: Multi refocused images in (a – e) and extended DoF in (f) (Ng et al., 2005).](image)
Subsequently, (Isaksen et al., 2000) demonstrated virtual refocusing from the LF for the first time. In recent studies, the original LF rendering graphics file was called a synthetic aperture photograph (Levoy and Hanrahan, 1996), it is suggested to take into consideration the interesting change of the inclined focal plane. For example, angle the film surface toward the main lens.

However, these refocus approaches have two limitations. First, large camera arrays are not suitable for capturing the necessary LF dataset and can be used for long-time scans associated with traditional photography or moving cameras for spontaneous shooting must be used. Second, because of inadequate inspecting of the virtual focal point gap (e.g., because of camera and camera hole), the outcomes tend to indicate high associating in the obscured area. The photo of LF will solve these two problems. As compared to the conventional camera, LF camera is also easy to use. Moreover, aliasing is significantly reduced with the optical design by the integration of the light that is passed through the aperture. Over all, the proposed technique should simulate the images after the actual shooting, tracks the refocused images, simulates the image super resolution techniques, and adds all refocused images to super resolved all-in-focus image.

2.9.1 The Auto Focus and Depth from Defocus

Part of the solution to the above limitations explained in the previous section were suggested by (Ng et al., 2005; Berkner and Shroff, 2011; Fiss et al., 2015). In addition, it has been studied by other scholars regarding the efforts to refocus from two images at variable depths. Generally, these techniques are put on the class of calculations of computer vision known as the defocus depths whose main task is to estimate the object’s depth based on the blur focused on different depths of two images. The depth of defocus will lead to an impressive
system for estimating the depth of video in real time. However, the system that refocuses from two images does not work well. Although it is possible to generate a reasonable image of the “virtual focal depth” that is close to “optical focal” plane for the input image (Fiss et al., 2015), the human factor rapidly increases, and it is possible to solve the more detailed problem. The basic issue is defocus, as it is a “low pass filter” which can effectively block the high frequency details of the distant optical focal plane (Berkner and Shroff, 2011). The basic principle of signal processing clearly shows it’s impractical to recuperate “attenuated high frequencies” by HR (Berkner and Shroff, 2011).

2.9.2 Depth of Field and Depth Estimation

According to (Georgiev and Lumbsdaine, 2009), the LF depth of field (DOF) is the array of focused multiple views appearing in an image. It is also known as the distance between the farthest and nearest blurred objects in a scene (Nagahara and Kuthirummal, 2008). It has been used widely in films and photography; also, it has been applied to virtual reality applications and image rendering. One of the fundamental limits in computational photography is the common between the size of aperture and the DOF. Despite recent attempts to solve this limitation, nevertheless, the limitation is from the design of the plenoptic camera. Consequently, (Bishop et al., 2009; Todt et al., 2008) show that the certain plane of focus is causing the low spatial resolution of the plenoptic camera. Therefore, next to this plane and the focusable range edges, the splat kernel size has to be increased.

In recent years, (Veeraraghavan et al., 2007) and (Levin et al., 2007) have controlled the defocus blur kernel properties by applying masks at the lens aperture. Their aim was to estimate the VPs arrangements to use the blur kernels by DoF corresponding. In addition, to deconvolve the reconstructed image and gain a full rendered image. Nevertheless, they
adopted weak layered views and their depth estimation recovery is not complete. One of the most common limitations in DOF is estimating the depth (Wanner et al., 2012). Below paragraph illustrates the limitations and work done previously in LF depth estimation.

The significant advantage of an LF camera for computer vision is the micro images or the various angles available in a single LF image which allow depth estimation. Furthermore, a depth map estimation mechanism for refocusing after capture is produced by Lytro and Raytrix software. Equally, (Tao et al., 2013) illustrated how various cues such as defocus and correspondence could be merged. However, there is a limited number of works that have clearly used occlusion in the past. The angular pixel matching to single spatial resolution denote various angles that match the same target in the view, is changed to the accurate depth such as the centre view depth (Jeon et al., 2015).

The work proposed (Perwass and Wietzke, 2012; Dansereau et al., 2004; Todt et al., 2008) used correspondence methods for depth estimation of the LF cameras. Furthermore, (Tao et al., 2013) complimented previous limitations of previous methods by combining correspondence as well as defocus cues in the 4D Epipolar Image EPI. Neither method explicitly models occlusions. On the other hand, a technique was proposed for the colour of the images in order to remove part of the occlusion (McCloskey, 2014). Furthermore, (Wanner and Goldluecke, 2012) applied a new approach to calculate the direction estimation of the feature pixels in 2D. Whereas (Yu et al., 2013) created a new approach to program the lines of the 3D array space to enhance the quality of reconstruction. Nevertheless, the two methods are weak to heavy occlusion: the estimation of the field is too random, and lines of the 3D ray space are divided into minor segments. A ‘fine-to-coarse’ method is proposed by (Tao et al., 2013) in order to enhance the homogeneous areas reconstruction using dense LF, while (Jeon
et al., 2015) adopted a phase-based interpolation algorithm to enhance the sub-pixel shift accuracy. Subsequently, the expected approach will be based on the algorithm in (Tao et al., 2013) which is based on LF cameras to increase depth estimation and implement a new method to enhance the LF occlusions. Also, (Tao et al., 2013) have adopted techniques for depth estimation, which intents at merging the shading cue. The novel method will be evaluated against existing algorithms such as (Bishop et al., 2009; Jeon et al., 2015; Wang et al., 2015) and results by Lytro (Ng et al., 2005).

2.9.3 Image Synthesis Light Field Refocusing

This section first discusses the recent approaches applied to LF refocused images of improving the hardware part by explaining the calculation of the refocus image in the LF. In the second half, it shows how to use the refocused image collection to widen the deepness of the field and make whole picture as clear as possible. In other words, enhancing the refocused images. The pixel’s ideal light regarding the digital refocus is a line of different rays that are converged to virtual conventional camera’s pixel focused at the required depth. The line of rays at position \( (l, k) \) on the pixel is specified by the derived equation by integration.

\[
E(\alpha, K)(l', k') = \frac{1}{a \in F^2} \iint_{LF} \left( i \left( 1 - \frac{1}{a} \right) + \frac{l'}{a}, j \left( 1 - \frac{1}{a} \right) + \frac{k'}{a}, i, j \right) di dj, \quad (2.16)
\]

Recall that LF is the depth of the \( i, j \) lens surface, \( \alpha \) represents virtual movie plane deepness with respect to \( F \). Whereas, \( (\alpha, K) \) denote the formation of image \( i, j \) parametric LF.

Integration can also be evaluated by applying various digital techniques like sampling of various values of totaling and \( i \) and \( j \). The ray tracing program described is used to evaluate the integrands of these different \( i \) and \( j \) samples. The idea is to keep track of light from
microlens array to photo sensor (Ng et al., 2005). This intersection occurs where light accumulates the dynamism in camera at the time of exposure & value of LF is estimated from the photo sensor value near that point. However, by integrating the relevant bright field linearity, a more efficient method is proposed by (Ng et al., 2005). Considering Eq. 2.16 an important observation that refocus is conceptually a sum of an extended version of the sub aperture and a shifted version becomes clear. This position is more cleared by the defining of the sub-aperture images at the bridge field LF’s lens position \((i, j)\). With the utilization of function \(L(i, j)\), sub gap picture is communicated in a way that pixel at position \((l, k)\) in the sub opening picture is \("L(i, j) F, k"\). In this documentation, it can be revamped as takes after.

\[
E(\alpha, K)(l', k') = \frac{1}{a^2F^2} \iiint LF(i, j) \left( i \left( 1 - \frac{1}{a} \right) + \frac{1}{a}, j \left( 1 - \frac{1}{a} \right) + \frac{1}{a}, i, j \right) di dj,
\]

Here, the partial aperture image \(L\), Shift \((l, k)\) expanded by \(\alpha\) factor. In other terms it can be stated that by increasing and moving the light field’s sub-aperture image, digital refocusing can be achieved. This technique is also used in the bright field synthetic aperture imaging that is obtained with the help of a camera array. By taking a closer look at the refocus magnification and offset of the current normal image. The expansion coefficient \(\alpha\) in Eq. 2.17 does not have an actual part when calculating the final image and can be simply ignored. This is because the expansion coefficients of all images are the same. Even if all the images are zoomed in and out, the final output resolution of the composite image does not change.

The true meaning of the amplification factor is related combined to fact that the virtual refocusing does not change viewpoint of subject. For most photographic lenses, the optical focus of lens near the object reduces the field of view of the object consequent to an increase of the object magnification. In contrast, since the mixed images possess fields of the view
related to original sub-aperture image, field of view remains constant regardless of the depth of focus selected for digital refocusing. Only the focus has been changed. In other words, the expansion coefficient of Eq. 2.17 represents an expansion of the domain of the depiction comparative to premises of the views that are obtained when the lens is optically focused to the desired depth. Regarding the characteristics of the changing focus deprived of alteration to the magnifications, digital refocusing is however functionally comparable to the telomere center lens, but basic reason is entirely opposite.

Offset \((i (1 - 1 / \alpha), j (1 - 1 / \alpha))\) of every VP in the expression two is the distance \((i, j)\) from the center of the sub-aperture to the lens. Figure 2.23 (a) visualizes the offset of three distinct digital film planes & adds two sub-aperture images \(l\) and \(k\) for the demonstration. Use two VPs to shift their concentration and make the offset effect easier to view with two overlapping edges instead of blurring. Figure 2.23 (b) corresponds to not refocusing and is shifted to zero for both sub-aperture images \(\alpha = 1\). Last two images indicate direction of offset depending that whether focus is approaching or approaching to align the features to the desired depth.

![Figure 2.23: Shift-and-add algorithm, demonstrated in addition to VPs (Ng et al., 2005).](image)
Minimum discretization of the current algorithm needs to be added in a $12 \times 12$ VP to original bright pitch. In many of the applications, quality of shooting results is best. Though, if the focal length is a distance away from the optical focal plane, the process may create undesirable step edges in the defocus area. This image shows the refocus of the image in the extreme foreground chain fence. In close-up of the out-of-focus area, edge artifacts can be seen as stripes. In the still images these artifacts are comparatively subtle, however these are always additionally prominent when you refocus the animation.

![Figure 2.24: Aliasing: Blurred the edges of the regions in “Under-Sampled Shift-And-Add Refocus” method (Ng et al., 2005).](image)

Due to the under sampling of the $i,j$ aperture, in the “numerical integration” of equations, these step edge artifacts exist as aliases. This problem occurs when the offsets of adjacent sub-aperture images differ for a plurality of pixels, therefore, single VP edges could not be merged easily with neighboring picture between the computation’s movements.
The arrangement is to super-example the gap plane and embed a sub-gap picture with a better determination than $12 \times 12$. The oversampling rate is chosen such that the minimum offset is less than one output pixel. Figure 2.24 (c2) shows an image without artifacts. Excess VP is incorporated through closest values of LF. The fourth order interpolation performance in the four-dimensional space is good.

When refocusing the focal plane, another artifact darkens the image boundary. This solar eclipse can be seen in Figure 2.24 (c1). The reason for this artifact is that some shifted sub-aperture images do not cover this boundary area. Another way to solve this problem is to not require light rays to estimate Eq. 2.17. These rays lie outside the physical boundary of the LF sensor are not ever calculated through camera & equivalent values of recorded LF function is zero however.

One arrangement of vignette is to standardize the estimation of the limit pixel by the extent of light really found in the recorded splendid field. For instance, the score is minimized with the most extreme boundary pixel. Adjust this score to the strength of the neighbor. Figure 2.24 (c1) calculates and eliminates the important dimming of Figure 2.24 (c2) using this normalization procedure.

2.10 Discussion and Summary

This chapter has addressed the comparison between the design of the plenoptic cameras and conventional cameras. Moreover, it provided the technical problems of the design since the invention of the camera. Besides, it has presented the history of the improvements the camera had starting from Daugher and ending by DSLR and the powerful design of the plenoptic camera.
As it is essential to increase the performance of LF images because of the limited spatial resolution, many attempts have been applied recently to solve this limitation. However, there is still a huge gap in their methods. For example, the research in (Bishop et al., 2009; Ng and Yau, 2005) applied a novel technique which only enhances the resolution by factor 4x4 of the sub-aperture images; it counts as a good attempt but still, it needs to improve to 8x8 and 16x16 factors compared to other studies such as in (Dansereau, 2014; Wanner and Golduecke, 2014). Moreover, the existing approaches of image SR such as (Park, 2003; Yang, 2010; Xu et al., 2012), have limitations including LR, smooth, blur, noise or high complexity. The first solution suggested, is applied to the classical 2D images. The outcome of this study has suggested a solution to these limitations by implementing novel approaches which will be explained in the next chapters. The second solution has been explained in this chapter, concentrated on the SR methods applied recently to the 4D LF images which photographed by the standard plenoptic camera. Nevertheless, the aim is to clarify the drawbacks of the existing SR techniques implemented, and how this advantage is taken so as to implement the proposed LFSR method. Third solution has addressed the limitations of the LF refocusing and how to apply the SR approaches in order to increase the all-in-focus images.

In conclusion, this chapter has first viewed the history of conventional and plenoptic cameras and explained the differences between them. Second, it has presented in detail the recent approaches of image SR in order to enhance the plenoptic images. The main benefit of this literature is to summaries the importance of the LF cameras and the needs of enhancing the LR images, as these cameras are emerging technology. Furthermore, this research addressed the changes of the camera design since it was invented until the current time. Cameras are in general needed in many places starting from public such as cinemas (film industry), roads,
shops and cities, and ending by government sectors specifically in army and private security companies. All these cameras need development methods to process their images specially the plenoptic camera as it suffers from LR which blocks the adoption of LF cameras extensive. Thus, the study concentrated on the methods applied to conventional images and LF images, precisely the classical approaches and plenoptic approaches. The classical image approaches have shown the merits and demerits of these techniques. In the same time, this study shown the pros and cons of the LFSR approaches. Most of the techniques explained in this chapter are used to super resolve the techniques of the LF images. Second, the chapter has discussed technical features the plenoptic camera has such as refocusing. The next three chapters show the contribution chapters proposed as well as explain the methodology with results and evaluation.
3. Hybrid Light Field Super-resolution Approach

This chapter presents a novel super-resolution (SR) approach to enhance images which suffer from low resolution (LR) when taken by light field (LF) cameras such like the Lytro camera (Ng et al., 2005). The proposed method is separated into two parts: the first part explains the initial combination of the method, which increases the resolution of multi-images that are captured by conventional cameras (2D images). The second part extends the method and is applied to LF images such as Lytro (4D images). The method is developed by taking advantage of the most recent classical SR approaches which combine multi-image super-resolution (MISR) and hybrid single image super-resolution (HYSISR). This combination is a novel approach, as it is based on a generation mask designed by a Laplacian pyramid. The masked image is segmented using an image segmentation approach to get the final super-resolved image. In addition, this image is super-resolved by super-resolution reconstruction (SRR) using sparse representation.
3.1 Introduction

Practical and theoretical limitations sometimes constrain the resolution that can be achieved by imaging devices. Image SR methods are created to tackle this limitation by fusing and acquiring several LR images which have the same scene for the purpose of creating high resolution (HR) images. Specifically, due to the nature of the plenoptic images neighbour views are generated at mutual shift that can be sub-pixel. That is the same as the starting point for the classical SR.

This chapter describes the methodology deployed in this research. Firstly, it shows how the resolution has been increased using the proposed hybrid imaging system method. Secondly, the results of the simulation are exposed to show the comparison between the results of the proposed method along with the achievements of other existing methods. After that, an evaluation of the performance is presented to measure the quality of the image resolution reconstruction. The next section identifies how the LF method is developed by addressing the problems of SR and plenoptic camera imaging. Subsequently, the LF method is developed based on the previous proposed successful methods, such as hybrid super-resolution (HYSR) and the LFSR (Bishop et al., 2009). Then, the chapter explains the simulation tools and the data collection methods. Finally, the performance evaluation is discussed with conclusions drawn.

3.2 Improving Resolution Using a Hybrid Imaging System

The initial algorithm is implemented using the proposed hybrid imaging method, which is a combination of two classical methods MISR (Park et al., 2003) and HYSISR (Xian et al., 2015). Moreover, the methods selected for SR are defined in a brief manner in the subsequent
subcategories. MISR, also called super-resolution reconstruction (SRR), is considered a significant method in various empirical cases. It takes multiple LR images, already captured or generated by other methods (out of the domain of SR); generates a HR image, then synthesizes single or multi HR images which have the same scene, such as video applications and medical and satellite imaging (Karimi et al., 2014; Macwan et al., 2014). However, a well-known method “single image super resolution” (SISR) is the technique used to capture a single image and applying interpolation methods such as bicubic interpolation to increase one single image from LR to HR. For more details the two methods, are explained in detail in the literature review sections 2.6.1 and 2.6.2.

3.3 Proposed Hybrid Super-Resolution (HYSR)

In this proposal, the two popular methods based on example-based learning HYSISR (Xian et al., 2015) and MISR (Park et al., 2003) are merged. Based on these recent approaches, the HYSISR method is a novel approach in hybrid SR. Furthermore, it allows merging of the output and input models. The method process is as follows. Firstly, one of the LR images is given as an input. Second, an HR image which matches the resolution of the target HR picture is created. Next, the pitches of the vertical and horizontal positions are superior, directed through matching gradients of the input image. Lastly, the calculated HR pitches are supply to uniform power function technique in (Xian et al., 2015) for the enhancement of the output resolution. Many tests prove that this approach exceeds the existing SISR methods both quantitatively and qualitatively. The next section shows how the proposed method is designed and modified.
Consequently, the overview of the proposed HYSR approach is presented in Figure 3.1. A similar approach was implemented by (Bätz et al., 2015; (Hu et al., 2012). Both approaches used hybrid imaging system techniques to increase the images’ resolution. The suggested method by (Hu et al., 2012) depends on a resolution setup that is mixed, where frames of non-key LR and frames of HR key are combined. In contrast, (Hu et al., 2012) method is based on a combination of MISR and SISR with a decision mask that is considered soft and dissimilar to the decision merging mask by (Hu et al., 2012) which is considered hard.

In comparison to the work of (Hu et al., 2012) and (Hu et al., 2012), the proposed approach is slightly better as it merges the HYSISR and MISR approaches to produce HYSR. Nevertheless, the input images are taken from sequence images, unlike (Hu et al., 2012) where they used a LR video sequence as an input. This section illustrates the HYSR approach.

Figure 3.1: Hybrid Super-Resolution framework

The HYSR proposed approach is clarified in Figure 3.1. An image that suffers from LR is taken from image sequences with frames $I_{LR}[i,j]$ and used as input. $[i,j]$ denotes the LR coordinates. Both $I_{MISR}[i,j]$ and $I_{HYSISR}[i,j]$ have to be calculated initially to super-resolve the current frame $I_{LR}[i,j]$ in a hybrid manner. While obtaining $I_{HYSISR}[i,j]$ only requires the current frame, computing $I_{MISR}[i,j]$ also utilises surrounding frames. The example-based
learning scheme, which is a SR method to enhance the resolution from a single image by (Xian et al., 2015) is used for HYSISR; conversely, an interpolation in (Park et al., 2003) is used for MISR, as described above.

\[ M_{xc}[i, j] \]  

(3.9)

where \( M \) represents the utilised frames quantity. According to (Hu et al., 2012), binary masks can be utilised with the intention of generating a soft decision merging mask, while combining two images. Therefore, a Laplacian pyramid which aims to detect edges is used to generate the soft decision merging mask, initially by creating a target image from \( I_{HYSISR} \) and also through generating a source image from \( I_{MISR} \). Furthermore, a test was carried out to determine the best images to perform around the edge region and to perform around the smooth region (See Appendix A). Moreover, it showed that the target image \( I_{MISR} \) performed better, therefore the mask is a binary image which identifies the pixels needed from the source \( I_{MISR} \) to accomplish the composite, as can be seen in Figure 3.2. After having computed both \( I_{HYSISR} \) and \( I_{MISR} \) from the binary masks \( M_{xc}[i, j] \) obtained from the MISR image, this mask is then upsampled using nearest-neighbour interpolation to the target resolution in order to obtain \( M_{xcu}[i, j] \). Initially, the decision mask

\[ H_{init}[i, j] \]  

(3.10)

is then designed as
Figure 3.2: Laplacian pyramid to extract the mask. Edge detector applied to the image at different scale. (a) MISR (Source image), (b) HYSISR (Target image), (c) MISR (Mask). The mask is a binary image which is generated by using the source image \( I_{MISR} \). The mask shows the pixels from the source which needs to composite with the target (See Appendix A).

Further secondary frames are able to underwrite additional statistics to the pixel when the initial decision mask's value is high. Therefore, a higher weight is received for the MISR result.
Fewer secondary frames are able to add to the pixel in the case of low value, and as a consequence, HYSISR ought to be favoured.

The final decision mask $H[i, j]$ is defined as

$$
H[i, j] = \left\{ \begin{array}{ll}
H_{\text{init}}[i, j], & \forall \in \mathcal{N} \setminus \mathcal{O} \\
\frac{N}{N}, & \forall \in \mathcal{N} \cap \mathcal{O} \\
0, & \text{otherwise}
\end{array} \right.
$$

(3.12)

The set $\mathcal{N}$ is comprised of each and every position, where the set $\mathcal{O}$ is comprised of the LR frame’s original sampling positions to be super-resolved. Hence, it is difficult to increase the spatial resolution by MISR, as positions with no-motion estimation can be part of the static background. However, HYSISR could be used for the purpose of enhancing the quality within these areas. Moreover, the size of the samples in the original positions are linked towards utilised frames $\mathcal{N}$’s quantity, so as to provide them with better influence for the fact that no motion compensation is used in order to acquire them. In addition to this, HYSISR may add synthetic high frequencies to decrease their values. The following equation is used to obtain the HYSR result.

$$
I_{\text{HYSTR}} = \left( \frac{H^2}{N^2} \right) I_{\text{MISR}} + \left( 1 - \frac{H^2}{N^2} \right) I_{\text{HYSISR}}
$$

(3.13)

The suggested hybrid method must serve as evidence of the notion as it makes input and augments the HYSR outcome’s impact. Different decision masks, as well as other groupings of MISR and SISR approaches are also possible and might further increase the quality and resolution of HYSR. Figure 3.3 illustrates in detail how the method is implemented.
Figure 3.3: Proposed Hybrid Super-resolution Framework.
3.4 Development of the Novel Method for Light Field Images

This section proposes a new LFSR method that can improve the quality resolution of plenoptic images, which combines the two classical SR methods explained above, including MISR and HYSISR. After this combination, the output light field hybrid super-resolution (LFHYSR) image is segmented into the background and foreground (i.e. objects of interest). Later, the sparse representation technique is applied to super-resolve the segmented image. As compared to the current light field super-resolution (LFSR) images, the images are obtained are of better quality than those obtained using the MISR (Park et al., 2003), HYSISR (Xian et al., 2015) and the epipolar image to calculate continuous disparity maps variational methods (Wanner and Goldluecke, 2014).

3.4.1 Light Field Super-Resolution Approaches

It is feasible for all the existing approaches for LFSR to be divided into two key classes, namely projection-based algorithms and variational optimization framework. The authors (Bishop et al, 2009; Bishop et al, 2012) enhanced image details by designing a variational Bayesian framework through increasing the resolution of the measured LF. This exploited prior knowledge about the scene and estimated both the LF and HR depth map, which extracts additional information from available data. Their method provides image formation by distinguishing the plenoptic camera point spread function (PSF) under Gaussian optics assumptions for a several-depth scene. In (Ng et al., 2005) the authors proposed a technique by using the epipolar plane image (EPI) to calculate continuous disparity maps and variational methods to calculate a novel super-resolved image structure of the LF. (Yu et al., 2013) performed Bayesian inference to achieve LFSR by using a Gaussian mixture model (GMM) for
LF patches. A simple and robust approach for LF images proposed by (Wang et al, 2016), is a projection-based method which does not require the parameters of the camera or setting, compared with existing LFSR methods.

Classical SR applied to 2D images has been focused on for a considerable time by previous researchers. The approaches implemented in (Hu et al., 2012) use hybrid SR techniques to enhance 2D images resolution. The proposed method in (Kim et al., 2010) is dependent upon a resolution setup that is mixed, where frames of non-key LR and frames of HR keys are combined. In contrast, the method in (Hu et al., 2012) is based on a combination of MISR and SISR with a decision mask that is considered soft, when compared to the decision merging mask of (Hu et al., 2012) which is considered hard. Despite their implementation of a successful SR method, their methods are operational only for the resolution enhancement factor up to 4. Moreover, none of the above approaches have used hybrid techniques to enhance plenoptic image resolution. In contrast, we propose a novel method that combines the two methods, MISR and HYSISR, to exploit the advantages of both. The combination is achieved as HYSISR methods are computationally fast, whereas MISR independently, can produce a better enhanced resolution via the available data taking neighbouring views into account. This combined technique is then applied to the LF images in order to enhance the viewpoint image (VPs) resolution. The method inherits image registration reliability and can reduce the artefacts in the SR images.

3.4.2 Proposed Light Field Hybrid Super-Resolution Approach

Initially, the method is implemented using the proposed method, that is HYSR. The LFSR method in (Bishop et al, 2009; Bishop et al, 2012) proposed an algorithm based on the Bayesian framework, assuming Lambertian textual priors in the image formation technique,
which depends on depth map estimation and light field HR. Furthermore, the algorithm that has been used is in accordance with the hypothesis that LF data are low-frequency signals. However, this method is weak, in sense that the LF disparity pattern is not clearly exploited. Consequently, the proposed method is implemented by enhancing the resolution of the VPs using the combination of MISR and HYSISR methods, which can largely reduce the artifacts and improve the precision of depth estimation. In particular, the enhancement of LF image resolution is achieved using the proposed HYSR algorithm in classical SR. A block-scheme for the proposed LFHYSR is presented in Figure 3.4.

Figure 3.4: A proposed hybrid light field super-resolution approach. Both MISR and HYSISR are combined and create a HYSR image. Compared to MISR and HYSISR the numbers in the picture are more readable in the LFHYSR result.
The LFHYSR approach is implemented for LF images to enhance the resolution of plenoptic images. The VPs are used as inputs and are written as

$$y^{(k)} \in \mathcal{R}^n \text{ with } n = n_x \times n_y \text{ and } k \in \mathcal{K} = [1, \ldots, K]$$  \hspace{1cm} (3.14)

where $k$ equals the number of the VPs. The purpose of this approach is to enhance the plenoptic image $x \in \mathcal{R}^N$, where $N = N_x \times N_y$. A SR method at this point is employed. First, $x$ is initialized. Eq. 3.15 shows the SR synthesis.

$$E_{SR}(x) = \sum_k \| y^{(k)} - I_{MISR} + I_{HYYSISR} \|$$  \hspace{1cm} (3.15)

Where $E_{SR}$ represents the energy function and $y^{(k)}$ represents the VPs, $I_{MISR}$ represents the multi-image super resolution method and $I_{HYYSISR}$ denotes the example-based model. To achieve $x^{(0)}$, one of the VPs corresponding to $y^{(k_e)}$ is generated with $I_{MISR}[\bar{m}, \bar{n}]$. Additionally, a single LR VP with a single HR image using $I_{HYYSISR}[\bar{m}, \bar{n}]$ has to be calculated to super-resolve the current frame $I_{LR}[\bar{m}, \bar{n}]$ in a hybrid manner, where $(\bar{m}, \bar{n})$ define the HR coordinates. The combination result of MISR and HYSISR images is then stored in the mask which was extracted using the Laplacian pyramid, as shown in Section 3.3, by means of the blending addition mode to blend the $I_{MISR}$ with $I_{HYYSISR}$, where the output image’s pixels are in linear combination to the input pixels. The composite image at level zero is to take pixels
by adding synthetic high frequencies. This process is defined by Eq. 3.16 to obtain the HYSR result.

\[
I_{HYSR} = (1 - H) I_{HYSIR} + H I_{MISR}
\]

(3.16)

where \( H \) represents the Gaussian for the mask generated from the source image \( I_{MISR} \), which is the blending ratio that determines the influence of each input image in the output. The ideal is to take mask pixels from the source image \( I_{MISR} \) and the other pixels from the target \( I_{HYSIR} \). After adding pixels, it will gain a HR image \( I_{HYSR} \).

After blending the \( I_{MISR} \) with \( I_{HYSIR} \), image segmentation is applied to the output HYSR image, as it shows an important role in defining an object from its background and is used to partition an image into meaningful regions and enhance the edges. As shown in Figure 3.4, the output HYSR image is segmented into the background and foreground (objects of interest) by way of the image segmentation method. The HYSR image segmentation is performed using the technique presented in (Graf et al., 2011). This technique is separated into several steps. Initially, the use of Gaussian Kernel to recognize the sharpness of pixels, is done by convolving the input image. After that, the image is subtracted from the blurred image, as a result the score to recognize region sharp pixels is provided. Thresholding comes next, where clustering takes place to threshold the value score and the pixel values by DBSCAN (Density Based Spatial Clustering of Applications with Noise). During this process, all isolated areas of high frequency are treated as noise and all big clusters are additionally run. While running image processing, a sharp noisy region is recognized. Regions which are added to the noise are overlooked again, on the condition that the number of pixels is not as much as half the size of the biggest cluster.
In this step, the clusters are used as input, which is created by DBSCAN. The input clusters are different single regions of interest that must be changed into a single attached region.

An estimated binary mask $H$ is a single attached region, which can be generated by joining the clusters by morphological filtering. In this step, due to colour similarity in $I$, the clustered pixels must be classified from the estimated mask $H$. A similar colour to all neighbours to $p$ is continuously formed and every pixel ($p(x, y)$) in $H$ is compiled. The outcome set of regions creates $R_{\text{color}}$.

In the final step, the score for each region is calculated $\in R_{\text{color}}$. The greater score value is distributed, which is bordered by regions of large scores. After this, moving the low score regions are moved to update the score in the neighbouring regions. At this point, any necessary repetition of deletion is applied to complete the high score of the updated region.

The HYSR image now segmented, the sparse representation technique is applied to super-resolve the segmented image. The energy function $E_{SR}(x)$ is as expressed in Eq. 3.17, where the term is a classical observation model obtained from the combination of the classical SR approaches mentioned above.

$$E_{SR}(x) = \frac{1}{2}\sum_{k \in \mathcal{K}} \| I_{\text{HYSR}} \|$$

(3.17)

The gain of the proposed technique lies in the fact that no method has joined MISR and HYSISR and applied them to plenoptic images. The advantages have become significant for SR image approaches.
**Algorithm: Hybrid Approach for SR**

**Input:** Viewpoint images input $y^{(k)}$ extracted from the raw image

1. To increase the VP’s quality, do the following
   
   a) For method one (MISR) $I_{MISR}$
   
   b) For method two (HYSISR) $I_{HYSISR}$

2. To merge two images, do the following
   
   a) Add pixel values of one layer to the other $I_{HYSR}$

3. To extract and segment the mask of object $I_{HYSR}$, do the following
   
   a) Identify pixels in sharp region for each pixel in $I_{HYSR}$
   
   b) Cluster generations from pixels using DBSCAN
   
   c) Link all clusters to one contiguous region
   
   d) For similar colours to all neighbours, create sets of regions in each pixel in $I_{HYSR}$
   
   e) For each region in $R_{color}$, update relevance score
   
   f) Eliminate noise to generate a segmented image

4. Apply SRR to the segmented image

**Output:** SR Image $x$
3.5 Evaluation Process Techniques

Evaluation requires quality metrics of the reliable image for the comparison of the performance of several SR methods. This is performed after successful development of the SR algorithms. Methods and techniques are evaluated both subjectively and objectively using the “peak signal-to-noise ratio” (PSNR), “structural similarity index (SSIM)”, CIEDE2000 as well as S-CIELAB superiority metrics. The performance evaluation outcomes are compared to (Bishop et al, 2009; Bishop et al, 2012) and bicubic interpolation numerically and objectively utilising the SSIM and PSNR.

The following singular indexes are used to evaluate the results: 1) “peak signal-to-noise ratio”, and 2) “structural similarity index measure” (Rwang et al., 2004). This is to evaluate the variance amongst two images, measure the reconstructed image’s quality and the similarity level to the original HR image.

1) The PSNR measures reconstructed image quality and compare it with the original image using a logarithmic scale. In that, MSE is used for two m x n metrics that represent the imageries I and K as well as equate the pictures. The PSNR can be expressed as

\[
PSNR = 20 \log_{10} \left( \frac{MAX}{\sqrt{MSE}} \right)
\]  

(3.18)

2) The SSIM contrasts and compares covariance and variance in the middle of the two images and measures the structure of an image. The SSIM (Rwang et al., 2004) can be expressed as

\[
SSIM(X, Y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)}
\]

(3.19)
The quality assessment of the proposed method is done by evaluating several images created by various bicubic interpolation and SR methods. The image quality is equalled to Bishops metric for SR methods evaluation. The performance evaluation is implemented using several standard test images, as well as readily accessible applications of SR algorithms within the Matlab platform utilised by the researchers of image processing. Thus, the outcomes could be effortlessly replicated. Another evaluation approach that will be used, if necessary is measuring the error of the image with registration freedom. It is possible to learn the registration for each of the K LR images to the SR frame when the image SR is performed in a the simultaneous setting (Zitova abd Flusser, 2003).

Figure 3.5: An illustration of gauge freedom in geometry. (a) Presents the three images generated. (b) Displays three same images registered as an SR image frame. For this reason, the frame at the centre has been skewed to the other two images, so the worst-case. (c) Shows the registration of the image.

For this example, the related registrations that have been used are the same as the true registrations, though their relationship to the correct scene frame has been altered by the transformation of the perspective. The variation can be accounted for in one homograph
within the ground frame truth and the SR image frame. For this reason, this is a suitable registration to use.

### 3.6 Tools and Data Collection

The main programming environment that has been used to implement the code is Matlab-15b since it supplies an efficient graphical user interface that implements several image reconstruction and registration algorithms for SR imaging. In addition, in the department, the plenoptic camera lab is used to capture the LF images as well as mathematical derivation and lab equipment for capturing data. The data are collected by taking some available images from the internet. Some images are captured by the plenoptic camera in (Hahne, 2016).

### 3.7 HYSR Approach Simulation and Results

#### 3.7.1 HYSR Results

The proposed HYSR is compared to other methods: SISR, MISR and HYSISR. Each test was implemented upon the three existing methods exemplified previously. The 15 frames of the dataset ‘Boy’ has a spatial resolution of 1024 x 1024 pixels. For simulation purposes, comparing the MISR method, the LR image sequence was despoiled through a 256 x 256 sized Gaussian blur and resampled through a factor of 2 and 4 within every aspect for assessing the enhancement of the resolution. Figure 3.6 shows the visual results of the ‘Boy’ image as it first shows the ground truth\(^3\) image with close-up images. After that, the visual comparison results are presented.

\(^3\) In this thesis, ground truth and reference image are used interchangeably.
Figure 3.6: The ground truth image used to test the HYSR method shown in (a). Whereas, close-up images of the ground truth image are shown in image (b) and image (c).
Figure 3.7: Comparison with SISR, HYSISR and MISR of an SR result on the ‘Boy’ image (magnified x2 and x4). The proposed approach reconstructs the ear and car contours with fewer artefacts, unlike other techniques which suffer from aliasing artefacts, jaggies and blurring. Finally, the results outperform the existing approaches in all datasets.
The implemented HYSR method is evaluated by other methods of resolution enhancement and assessed with reference to the visual impression and objective quality. For the purpose of measuring the quality of objective, the luminance PSNR is utilised. SISR methods do not require motion estimation and the resolution enhancement has fundamental limitation; however, they are considered robust. Yet, the HYSISR method is computationally faster. MISR, on the other hand, through taking statistics from neighbouring frames into consideration, can achieve better enhancement resolution.

**Table 3.1:** Enhancement resolution by factors 2 and 4 of the average PSNR outcomes in dB for $N = 5$

<table>
<thead>
<tr>
<th>Boy</th>
<th>x 2</th>
<th>x 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>SISR</td>
<td>HYSISR</td>
</tr>
<tr>
<td>PSNR [dB]</td>
<td>36.48</td>
<td>37.39</td>
</tr>
<tr>
<td>Average gain over SISR</td>
<td>_</td>
<td>0.91</td>
</tr>
</tbody>
</table>
Figure 3.8: Average PSNRγ gain over SISR for a varying quantity of utilized frames. Effects assumed an upscaling influence of 2 and 4. The proposed HYSR approach is denoted by the solid lines.

3.7.2 HYSR Performance Evaluation

After the successful development of the SR algorithm, previous and proposed methods are evaluated using a range of techniques such as PSNR and SSIM, and the results are analyzed, both subjectively and objectively. The results of the proposed HYSR method are compared to SISR (Kim et al., 2010), MISR (Park et al., 2003) and HYSISR (Xian et al., 2015). Figure 3.8 presents the average of the PSNR gain over all frames sequences and all competitive methods associated to SISR for a variable quantity of frames $N$, as shown in Table 3.1 for upscaling factors of 2 and 4. For proof of concept, the proposed algorithm was verified for a variable quantity of secondary frames as well as upscaling influences of 2 and 4. Overall, the outcome
of this evaluation shows that the HYSR technique produced a better super-resolved image. Also, the proposed method achieves an overall average gain of up to 1.79 and 1.71 dB in luminance PSNR for factors upscaled by 2 and 4. The performance drops sharply because of the regression of number of frames when increasing up to 7 or 9, as well as corresponding differences in pre-processing.

3.8 LFHYSR Approach Simulation and Results

The proposed method is a working application for images of conventional and plenoptic cameras. In this study, we initially apply the method to the multi-frames “boy” image which is captured using conventional camera from different angles as shown in section 3.7.1. For plenoptic images, the modification is necessary since the plenoptic implementation has a mechanism to follow. The implemented code is modified to process the input as Raw images to extract the 2D images, after that the method is applied.

This section presents the whole process of the simulation applied to the LFHYSR approach. Starting from the mechanism of the LF camera moving to an algorithm design and ending with the comparison of the proposed approach with existing LFSR methods.

3.8.1 Raw Data Analysis

The plenoptic proposed implementation has a mechanism to follow. The plenoptic camera captures raw (unprocessed) images from different VPs. Then, the VPs images are extracted from the raw image by a pixel at a very similar position from every microlens and positions them all. The VPs are able to be observed as the rectangular images set which perspectives are organised upon the surface conforming to the LF camera’s aperture. Due to the LR of VPs, the SRR is applied to enhance these images. This section illustrates how raw data are analysed
out of the plenoptic camera and explains the calibration steps between the image sensor and microlens array.

The raw image is placed in a file format ‘.lfp’, after it is taken, and the parameters of the camera, for example the focal length, are also held in this .lfp file format. The file contains the RGB coloured raw image; moreover, the resolution of the raw image has 2529 x 1683 pixels and 9-bits per pixel are stored. It is worth noting that the LF camera microlens has the shape of a hexagonal arrangement, which is the reason the shape of the raw image is hexagonal when it is magnified, as shown in Figure 3.9. Additional light rays are allowed to be captured because of a smaller gap between microlenses. For instance, the diameter for each microlens is around 9 pixels and physically each microlens size is $1.4 \times 10^{-5} m$ (Cho et al., 2013). Therefore, if the images’ dimension is separated by the microlens size, the image resolution of each VP image is 281 x 187.

Figure 3.9: Plenoptic raw image and magnified images.
In more detail, 91 VPs are extracted from a raw image. To convert the raw image file into a LF image, the raw image must be calibrated (Hahne, 2016). The calibration purposes to identify each micro-lens sub-image centre point location and reorganise them for better resampling. A white image is captured to calibrate the raw image and all captured images should be homogeneous and white in colour. The white images are taken several times and the average white image is used for calibration. To decrease the sensor noise effect during calibration, gamma correction is applied for each individual capture to accurately measure the intensity of where the value of gamma is placed in the .lf file. Subsequently, because the colour of the image taken is white, the values of the RGB colour channels must match and are used to demosaic the correct image colour. Furthermore, to stretch the intensity range, the RGB is transformed into a grayscale image, thus making it easier to process the image in the following phases.

3.8.2 Experimental Outcomes and Performance Evaluation

In this section, extraction of the VPs is illustrated, and the performance is evaluated. It is the first process applied after extracting images from the plenoptic camera. Figure 3.10 and Figure 3.11 show the VPs after they have been generated and extracted from the raw image. All LF images $L(u, v, s, t)$ used in the proposed method are reconstructed with size 9 x 9, using only the pixels that surround them. The resolution of the raw image results is rendered for ‘Numbers’ and ‘Toy’ at 2529 x 1683 pixels, however Lytro rendered their images at 1080 x 1080 pixels (Ng et al., 2005). For this detailed comparison, Lytro used bicubic interpolation to up-sample the results, whereas the SRR technique is used in the proposed method to enhance the segmented image.
Consequently, the 4D function of the LF, can be officially denoted as $\Pi \times \Omega \rightarrow \mathbb{R}$, $(u, v, s, t) \rightarrow L(u, v, s, t)$ in which $\Pi$ remains consistent to their angular allocation, marked by $(s, t)$; while the $\Omega$ denotes the allocation of spatial light rays, marked by $(u, v)$. Consider the LF as a matrix of views, also called VP images, simplifying the way of visualizing. A 2D slice of the LF is represented by each VP in the spatial dimensions $(u, v)$, as shown in Figure 3.10. Besides, Epipolar Plane Images (EPI) are a popular representation of a LF, acquired by settling one angular dimension and one spatial $(su \text{ or } tv)$.

![Figure 3.10: Example of the representation of LF: View Point Images (left); EPI (right)](image)

The technique that has been followed by the proposed method is to extract the VP, rather than EPI, enables more LR images to be chosen. Microlenses are organised in a hexagonal grid (Hahne, 2016). The number of microlenses is denoted by $l = l_x \times l_y$, where $x$ and $y$ are the
horizontal and vertical of LF image directions. A set of pixels captured by each microlens is called a ‘sub-image’. Pixels from each sub-image are gathered to create a VP image, which is located in the same position with respect to the centre of the sub-image. Consequently, in each sub-image, the number of VPs is equivalent to the pixel numbers in each sub-image. The number of pixels of a VP \( n_x \times n_y \) is equal to the number of microlenses, that is: \( n_x = l_x \) and \( n_y = l_y \). The VPs extraction results, as displayed in Figure 3.11, are the ‘Numbers’ image captured at the University of Bedfordshire by (Hahne, 2016). However, the example above is a ‘Lego’ image taken from Stanford University (Vaibhav, 2008).

![Figure 3.11](image)

**Figure 3.11**: Above: Raw image captured by the plenoptic camera (Hahne, 2016). Below: 9 x 9 viewpoints, rearranged as a collection of several views.
The RGB is converted into a grayscale image and the intensity range is stretched to process the image in advanced phases.

![Grayscale raw image and 9x9 VPs](image)

**Figure 3.12**: Above: Grayscale raw image. Below: Grayscale 9x9 VPs.

In the experiment, as mentioned previously, the raw images $L(s, t, u, v)$ are captured by the plenoptic camera and are reconstructed with size $9 \times 9$ sub-aperture images, as shown in Figure 3.12.

Consequently, to compute the performance quality of the novel approach and other exiting approaches, a ground truth image is generated by creating LR images from the VPs such that...
each VP image has $140 \times 93$ image resolution. Moreover, these images were interpolated using the linear interpolation technique to reach the resolution target. Each of the VP resolutions was set to $1124 \times 744$. VPs have been converted to a grayscale image as it is easier to apply the interpolation technique to create the ground truth image (reference image) of the reference VP. The size of VP is $281 \times 186$ pixels. The resolution is enhanced by factor 4x4, which means the interpolated image is increased to $1124 \times 744$ pixels.

Figure 3.13: The result from interpolation (fill factor 1/9). **Left:** The ground truth reconstructed VP image (one view of 81 LR frames) **Right:** The interpolated VP image.

Figure 3.14: **Left:** A grayscale VP image cropped from a VP image. **Right:** Interplated cropped VP image which is used as a reference image.
The proposed LFHYSR method has competed with the existing resolution enhancement approaches and it was compared to bicubic interpolation, MISR (Park et al., 2003), HYSISR (Xian et al, 2015) and (Wanner and Goldluecke, 2014). The images presented in Figures 3.15, 3.16, 3.17 and 3.18 show the four super resolved central view images ‘Numbers’ ‘Toy’ ‘Truck’ and ‘Flowers’ which were selected to show the comparison results for the SR. Two datasets ‘Numbers ‘and ‘Toy’ where captured by our plenoptic camera (Hahne, 2016) and the ‘Truck’ and ‘Flowers’ were taken by Stanford University (Vaibhav, 2008). Any datasets can be chosen to be tested by the proposed approach, but the selected datasets are higher and better quality compared to many plenoptic images and more specifically is that the most recent studies have used these datasets. The performance and the results show that the proposed method outperforms existing comparative methods in terms of resolution and independent picture quality. The proposed approach is evaluated using a range of results analysed both objectively and subjectively, using the structural quality matrices PSNR and SSIM. To show better visualization the images are cropped. Furthermore, the evaluation results on the proposed LFHYSR approach are shown in Table 3.2. The bicubic method’s performance drops sharply because of the regression of VPs and corresponding differences in pre-processing, even more so than MISR and HYSISR approaches. The proposed approach can achieve an average in PSNR and SSIM of up to 1.06 dB and 0.0107 for the Numbers’ image, 0.97 dB and 0.0028 for the ‘Toy’ image, 1.16 dB and 0.0135 for the ‘Truck’ image and 1.08 dB and 0.0111 for the ‘Flower’ image.
Figure 3.15: Comparison results for super-resolution ‘Numbers’ image. The left column is the full picture. The right column is the close-up. (a) Reference (b) Bicubic method (c) HYSISR (Xian et al., 2015) (d) MISR (Park et al., 2003) (e) (Wanner and Goldluecke, 2014) (f) LFHYSR the proposed method
Figure 3.16: Comparison results for super-resolution Toy image. The left column is the full picture. The right column is the close-up. (a) Reference (b) Bicubic method (c) HYSISR (Xian et al., 2015) (d) MISR (Park et al., 2003) (e) (Wanner and Goldluecke, 2014) (f) LFHYSR the proposed method
Figure 3.1: Comparison results for SR Numbers image. The left column is the full picture. The right column is the close-up. (a) Reference (b) bicubic method (c) HYSISR (Xian et al., 2015) (d) MISR (Park et al., 2003) (e) (Wanner and Goldluecke, 2014) (f) LFHYSR the proposed method.
Figure 3.18: Comparison results for super-resolution Flower image. The left column is the full picture. The right column is the close-up. (a) Reference (b) bicubic method (c) HYSISR (Xian et al., 2015) (d) MISR (Park et al., 2003) (e) (Wanner and Goldluecke, 2014) (f) LFHYSR the proposed method.
### Table 3.2: Average PSNR and SSIM results

<table>
<thead>
<tr>
<th>Methods</th>
<th>PSNR (dB)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Toy</td>
<td>Numbers</td>
<td>Truck</td>
<td>Flower</td>
<td></td>
</tr>
<tr>
<td>Bicubic</td>
<td>33.27</td>
<td>33.64</td>
<td>34.21</td>
<td>34.11</td>
<td></td>
</tr>
<tr>
<td>HYSISIR</td>
<td>35.27</td>
<td>35.44</td>
<td>35.65</td>
<td>35.44</td>
<td></td>
</tr>
<tr>
<td>MISR</td>
<td>35.48</td>
<td>35.65</td>
<td>35.82</td>
<td>35.76</td>
<td></td>
</tr>
<tr>
<td>Wanner and Goldluecke</td>
<td>35.91</td>
<td>36.12</td>
<td>36.61</td>
<td>36.30</td>
<td></td>
</tr>
<tr>
<td>LFHYSR (Proposed)</td>
<td><strong>36.45</strong></td>
<td><strong>36.71</strong></td>
<td><strong>36.98</strong></td>
<td><strong>36.84</strong></td>
<td></td>
</tr>
<tr>
<td>Average gain over MISR</td>
<td>0.97</td>
<td>1.06</td>
<td>1.16</td>
<td>1.08</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>SSIM</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Toy</td>
<td>Numbers</td>
<td>Truck</td>
<td>Flower</td>
<td></td>
</tr>
<tr>
<td>Bicubic</td>
<td>0.9184</td>
<td>0.9341</td>
<td>0.9372</td>
<td>0.9332</td>
<td></td>
</tr>
<tr>
<td>HYSISIR</td>
<td>0.9398</td>
<td>0.9370</td>
<td>0.9396</td>
<td>0.9351</td>
<td></td>
</tr>
<tr>
<td>MISR</td>
<td>0.9409</td>
<td>0.9391</td>
<td>0.9425</td>
<td>0.9412</td>
<td></td>
</tr>
<tr>
<td>Wanner and Goldluecke</td>
<td>0.9417</td>
<td>0.9421</td>
<td>0.9480</td>
<td>0.9456</td>
<td></td>
</tr>
<tr>
<td>LFHYSR (Proposed)</td>
<td><strong>0.9437</strong></td>
<td><strong>0.9498</strong></td>
<td><strong>0.9560</strong></td>
<td><strong>0.9523</strong></td>
<td></td>
</tr>
<tr>
<td>Average gain over MISR</td>
<td>0.0028</td>
<td>0.0107</td>
<td>0.0135</td>
<td>0.0111</td>
<td></td>
</tr>
</tbody>
</table>
3.9 Summary

Overall, this chapter has proposed a novel SR approach. First, it has presented a new classical approach which can be applied for multi LR images in order to obtain a HR image. Second, it has shown a new method for plenoptic images which is capable of enhancing the resolution of the images. This is achieved by combining the classical approaches of MISR and HYSISR techniques. After combining these techniques, image segmentation is applied to the output HYSR; then a sparse representation technique is applied to super-resolve the segmented image. The results competed with existing resolution enhancement approaches and were assessed with respect to visual impression and objective quality. Nevertheless, the input images were taken from sequence images (plenoptic camera), unlike (Hu et al., 2012) where they used a LR video sequence as an input. The proposed method performance is measured by both the PSNR and SSIM. The final result shows that the proposed approach exceeds other methods with an average achievement over MISR in PSNR of up to 0.97, 1.06, 1.16 and 1.08 dB as well as 0.0028, 0.0107, 0.0135 and 0.0111 for SSIM for the ‘Toy’, ‘Numbers’, ‘Truck’ and ‘Flower’ images. Finally, the outcome of this evaluation shows that the LFHYSR technique acquired a better super-resolved image.

The next chapter presents a development of the proposed method which uses different technique applied to the viewpoints (VPs) by exploiting disparity information as well as the next chapter shows the experimental results and the performance evaluation.
This chapter proposes an extended approach of block matching (BM) for plenoptic images in order to enhance the resolution of the light field (LF) images by reducing the computational complexity. The BM approaches that have been developed for LF images, were applied to enhance the depth map. However, the proposed method aims to use BM to the viewpoints (VPs) by exploiting disparity information. This method is referred to as light field block matching super-resolution (LFBMSR). Moreover, the method forms disparity compensated estimated blocks from the central VPs by taking the advantage of the LF 4D nature. The block in the centre of the VP will compare the surrounding blocks and choose the best matching block. Furthermore, the average of this estimation is weighted after the final block is obtained.
to form a 4D transform domain, in order to obtain the final image. In addition, the directionally adaptive interpolation method in (Velisavljevic, 2008) was implemented to enhance the resolution of the VPs by 4x4 factor, before applying the LFBMSR in order to simplify the LR views for better processing. As a result, the proposed LFBMSR approach outperformed the result from the proposed approach in (Farag et al., 2017) and the related most recent approaches. Furthermore, it provides super resolved HR images, both subjectively and numerically. It exceeds other competitive methods because of the interpolation method used and exploited the disparity information applied contributing more. Consequently, the test results demonstrate that the proposed approach exceeds other methods in terms of computational complexity and the two-objective metrics of image quality PSNR and SSIM.

4.1 Introduction

Block matching is a popular technique in the area of computational photography due to its simplicity and effectiveness for software programming, specifically in video coding. In most cases, BM is used for video compression and many other applications, such as image denoising. It has been assumed that the movement of pixels within a defined region of the current frame can be modelled by considering the best similar match block in another frame. Since it is important to use BM in image and video coding, it has been found useful to exploit this technique in the images of plenoptic cameras. Since the main drawback of the current LF camera is the low spatial resolution, it makes it difficult to utilise them in applications where high-resolution (HR) is key; for example, in movie production. BM is necessary to search for
the neighbouring best matched blocks to the central view’s blocks in order to gain a SR image, by exploiting the disparity information of the VPs.

In this chapter, section 4.2 explains the steps of disparity information extraction which were used to find the distance of the depth between the views. Section 4.3 presents the proposed LFBMSR approach. The approach is derived using BM2D in (Lebrun, 2012) and BM3D in (Dabov et al., 2007). In section 4.4 and 4.5 the performance evaluation and the outcomes are displayed. Finally, the chapter summary is given in section 4.6. The recent studies of LFBM approaches applied to classical and LF images are explained in the literature review Section 2.7.

4.2 Block Matching with Depth Refinement

The disparity extraction is the process of finding the distance between the views by using two stereo cameras, also known as the ‘binocular stereo system’ (Hahne, 2016). Nowadays such a system is considered redundant technology. The advantage goes to the plenoptic camera which is capable of performing on the sequence by displacement analysis in order to determine depth. A research of the relationship between the baseline and the accuracy of estimation of the depth in a stereo system is explained in (Grimson, 1981). The research showed that the ability to estimate the depth is considered one of the most important features of LF cameras. However, microlens array-based LF cameras have low accuracy due to the narrow baseline.

Recently, BM approaches have been researched in a number of studies, such as (Danieyan et al., 2008; Nasrollahi et al., 2014; Licheng et al., 2016; Marenzi et al., 2017; Maggioni et al., 2012). The authors applied the BM approach to LF images as they lay to VPs freely. Their
studies have specifically shown the usage of BM techniques in 3D images. Regarding exploiting BM for denoising, the EPIs were discussed by (Baker et al., 2013) which initially denoised the angular and spatial images. For example, \((y, t)\) plane represent the angular distribution and \((x, s)\) denote the spatial distribution. It follows, this initial measure utilises the corresponding EPIs \((y, t)\) plane as it is significant of the angular distribution. Despite this fact, such approaches only consider features of the 2D from the 4D LF. Nonetheless, the approaches illustrated exclusive performance, however the structure of the 4D LF neglects entirely the technique of BM3D, as well as bypassing video block matching 4D (VBM4D) (Dabov et al., 2007).

Consequently, the existing methods generated to extract the disparity information are numerous, such as BM and stereo matching, as highlighted in section 4.2.2. In contrast to the highlighted methods, the proposed LFBMSR approach in this chapter extends the technique of LFHYSR. The method in (Sabter et al., 2015) has been adopted and used to exploit the disparities information of the views. Furthermore, the method was chosen after a small number of quality comparisons between various BM depth estimation methods, such as (Dabov et al., 2007) and (VBM4D) (See Appendix B). The method is demonstrated by completely taking into account the 4D LF structure and its ability to exceed the recent competitive methods. The process is carried out by generating selected views from the available 9 x 9 views. For each of these views the disparity is estimated. Each pixel in the image corresponds to the pixel in the depth image, then the intensity is known of the disparity of each particular pixel in this scene, for each view. Moreover, the registration process is redesigned to simplify this to a classical super-resolution so as to obtain a higher result at the end.
Figure 4.1 illustrates the process of exploiting disparity information from the generated multi-LR views extracted from the raw image.

**Gathering of differently views by changing the assumed depth, i.e. in (0, -4) represent the VPs location (0) is the central VP and (-4) is the far-left image.**

**Multi-viewpoint images (9x9)**

**Block Matching**

**Figure 4.1**: Disparity information extraction which is used to find the distance of the depth between the views, from left to right: the dark images show the processed VPs that are in the far left and right, the rest shows readable disparity maps.
4.3 Proposed Light Field Block Matching Super-Resolution

The proposed LFBMSR is a novel approach which is a combination of several existing super and interpolation methods as listed below:

1) The directional adaptive interpolation is applied to the VPs to increase the image size by a factor of 4x4.

2) The iteration of BM over VPs is applied to find the best similar block in $S$. Specifically, one of the VPs is selected as a reference of the central VPs of the LF.

3) Consideration is given of the neighboring VPs found in a (so-called) “angular search window” of size $n_u \times n_u$, and all of these VPs are run together.

4) The iteration is done over the 2D blocks of the reference VP, until it finds the optimal similar block that matches the reference block.

5) 4D blocks are then formed of the reference VP, exploiting the information of the neighboring VPs by constructing 2D blocks. After that, the rest of the remaining pixels in neighboring VPs need to be processed for a second iteration.

6) After fully processing the neighboring VPs with current reference, consideration is given to the remaining VPs to generate a new reference VP.

After the LFBMSR is complete, the proposed LFHYSR implemented in chapter three is used to conclude the process. This combines the BM (HYSISR) and BM (MISR) classical approaches. The combination of MISR and HYSISR images is then stored in the mask object using the blending addition mode, where the output image’s pixels are in linear combination to the input pixels. The composite image at level zero is to take pixels by adding synthetic high frequencies. This process is defined by Eq. 4.1 to obtain the HYSR result.
\[ I_{\text{HYSR}} = (1 - H) I_{\text{HYSR}} + H I_{\text{MISR}} \]  

(4.1)

Where \( H \) represents the Gaussian for the mask generated from the source image \( I_{\text{MISR}} \). This is the blending ratio that determines the influence of each input image in the output. The ideal is to take mask pixels from the source image \( I_{\text{MISR}} \) and the other pixels from the target \( I_{\text{HYSR}} \). After adding pixels, it should result in a HR image \( I_{\text{HYSR}} \). Image segmentation is then applied to the formulated \( I_{\text{HYSR}} \) in order to segment the super-resolved image, as it shows a crucial role in defining an object from its background and is commonly used to partition an image, so that the edges will be enhanced. To obtain the final LFHYSR segmented image, SRR is processed which is implemented and intended to be the last step applied in of the proposed method. Figure 4.2 shows what has been explained in this section to clarify the process of the implemented algorithm.

**Figure 4.2:** A block diagram of the proposed LFBSR approach: the red box shows 3x3 VPs and it selects a central view as a reference to search similar blocks to the central blocks. The number of views is free to choose, in this work the VPs used are 3x3, 4x3, 4x4 and 5x5 VPs.
Since it is the explanation of the methodology sections, a full demonstration of the implementation process is explained in detail below after a brief overview of the proposed method above.

### 4.3.1 Directionally Adaptive Interpolation

Initially, the proposed method applied an interpolation approach, as shown in Figure 4.3, which is a directional, adaptive approach adopted from (Velisavljevic, 2008). The aim of using this method is to upscale the factors image by 4x4 of all VPs, as there is a fundamental limitation when increasing the resolution 8x8 or 16x16. The method changes the direction of the transform and results in the provision of proper matching among the transform, across segments and locally dominant directions. For more simplicity, the direction pairs of horizontal and vertical VPs are distributed for the segments, which are required to be smooth with the elimination of the deceptive dominant direction. Different methods could have been used to upscale the VPs images, however this approach has been chosen since it outperforms other interpolation methods based on both numeric and visual quality of interpolated images.

![Directional Adapting Interpolation](image)

**Figure 4.3**: Directional Adapting Interpolation, which applied to upscale the VPs resolution
4.3.2 Outline of the Proposed LFBMSR Approach

BM is utilized in the implemented technique to form groups of blocks. In addition, the input VPs used are divided to a fixed window size of square blocks (See Appendix B). Technically, the procedure of the proposed method is carried out as follows:

The surrounded input VP images with HR can be seen in Figure 4.4, they are processed to extract reference blocks from the $l_{ref}$ successively, for each such block search for these estimates. Blocks which are best matching the reference one, are stacked together to create a group.

The group blocks are processed collaboratively, so that it can obtain the 2D estimate from the original places of group blocks. However, the estimated blocks obtained could overlap when all the reference blocks are processed and consequently, it is necessary to have a number of estimates for each pixel. Thus, to obtain estimates of the complete image, these estimates are aggregated.

Two varied steps are applied throughout the implemented method process. The first one is basic estimation and second, is final estimation. The whole process of the method that is shown in Figure 4.4, is applied as follows:

1) Basic estimation: Apply to the reference VP selected for each block to estimate wisely all blocks of the neighbouring VPs to be constructed. Next, search for the best similar block matches to the current constructed ones and stake them as group. Following on, apply blocks construction hard-thresholding which uses a 4D transform to the stacked group. After which, utilise transform coefficients through the hard-thresholding technique, attenuate the reference image by inverting the 4D transform of all grouped block estimations and return the blocks estimates back to their main locations. Then apply mapping to the grouped blocks
which computes the estimate of the basic step applied to the reference image, by measuring the block-wise estimates overlapped averages.

2) Final estimation: Use the basic estimation to implement an enhanced hard thresholding blocks construction group. First, group each block by block-wise estimates. Then, to search for the location of the blocks that are similar to the currently processed one, use the basic estimation within BM.

Consequently, the locations are formed mainly by two groups, one is formed from the basic estimate and one from the reference image. Furthermore, in order to apply a 4D transform on both groups, block construction processing should be used. Additionally, block construction processing performs the 4D transform process to utilise the spectrum as a basic estimation process of the main spectrum.

Every group of blocks should be predicted by implementing the reverse 4D change on the separated coefficients and the predicted blocks’ position restored to their original position.

The last step in this estimation process is known as aggregation. In this process, it uses a weighted average. From the true image it computes a final prediction by aggregating overall acquired local estimation. There are mainly two motivations behind choosing this second method, first of all, it utilises the basic estimation, rather than the reference image which enhances the grouping process in block matching. Secondly, the pilot signal processing utilises the basic estimation process, which is more powerful and robust, in comparison to the simple hard-thresholding 4D range of the reference image.

To illustrate the key contribution between the proposed LFBMSR technique as well as existing techniques, such as BM3D by (Dabov et al., 2007; Lebrun, 2012), their method applied BM3D for LF denoising, rather than SR. However, the main aspect is that the technique proposed
uses BM for enhancing the resolution of LF images. According the literature, the novel approach is new as it combines BM with LFSR methods. For instance, the two estimation steps applied, have assured the quality and robustness of the method by searching for the best similar blocks and discarding the unnecessary blocks, thus forming a group of comprehensive 4D transform.

4.3.2.1 Construction of the Blocks

The second step of the method is to generate block construction, as it is an essential step to BM. The proposed approach takes the advantage of the two steps used in BM3D: as explained in the section above, where at first, a basic estimation is obtained as (first iteration) by using the technique of hard-thresholding applied to coefficients transform of the 4D blocks. After, the basic estimation is exploited to find the best similar BM between all VPs, then the basic estimation is used to perform the domain of the 4D transform, as well as produce the final image estimate. A brief description is shown below to the proposed LFBMSR approach. For each block \( B \) of magnitude \( k \times k \) in the VP \( I_{ref} \) reference image, a 4D block is formed by the exploited \( I_{ref} \) redundancies and its neighboring

\[
\text{VPS}\{l_s,l_t\}, \in [1,n_a] \times [1,n_a]
\]

(4.2)

4D block is formed by looking for the 2D block closest to \( B \) in each neighboring VP, which can be accommodated to a disparity offset step by the method of BM. Formally, the 4D block of size

\[
n_a \times n_a \times k \times k
\]

(4.3)

is defined as
\[ \mathcal{B}_d(B) = \left\{ R_d^{s,t} : R_d^{s,t} = \arg \min d(B, R_d^{s,t}) \right\} \] (4.4)

Therefore,

\[ R_{s,t} \in FW_d^{s,t}, d(B, R_d^{s,t}) \leq \tau_d \} \] (4.5)

Where \( d(B, R) \) the normalized quadratic distance between blocks, \( FW_d^{s,t} \) represents the search window in \( I_{s,t} \) of magnitude \( n_d \times n_d \) centered on the position of \( B \). As well as \( \tau_d \) denotes the threshold for distance \( d \) based on which blocks are presumed comparable. Consequently, the idea behind threshold \( \tau_d \) is to discard the less matches blocks compared to the reference blocks in order to occlude robustly. Figure 4.4 shows the algorithm outline of the BMSR approach.

**Figure 4.4:** Search window of the LFBMSR algorithm outline. With the help of the search window, the VPs of each LF are processed in iteration by taking into consideration the reference VP alongside its surrounding VPs. 4D blocks are usually acquired when VPs are processed with angular search window, through the use of disparity compensated 2D blocks within the surrounding VPs with respect to a 2D block in the reference VP.
4.3.2.2 Block-Similarities

In this stage, the method looks in the reference VP for a number of blocks best similar to \( B \), determined by

\[
\mathcal{B}_{\text{sim}}(B) = \{ R_{\text{sim}} : d(B, R_{\text{sim}}) \leq \tau_{\text{sim}} \in FW_{\text{sim}} \} \tag{4.6}
\]

Where \( \tau_{\text{sim}} \) is the threshold distance for \( d \) under the estimated similar blocks, \( FW_{\text{sim}} \) and \( FW^{s,t}_{\text{d}} \) is a search window in \( I_{s,t} \) of magnitude \( n_d \times n_d \) centered on the position of \( B \). And \( \tau_d \) is the threshold distance for \( d \) under the estimated similar blocks.

4.3.2.3 Mapping and 4D Transform

After that, the method gets estimates for each 2D block belonging to a 4D block after the mapping is done, and thus, estimates each pixel in all blocks. Then the average of all estimates is weighted after the final block estimate is obtained. Below in Figure 4.5 is an example how the mapping is obtained to all blocks.

Figure 4.5: Mapping and 4D transform
4.3.3 Light Field Hybrid Super-Resolution Approach

The explanation details of the proposed LFHYSR approach was discussed in chapter three, however, it is briefly presented in this section as it is required for the process of the final resolution image to be obtained for the LFBMSR. The algorithm steps are as follows:

1) The mask is a binary image which is generated from the source image MISR.
   - The merge of multi-image super-resolution (MISR) and hybrid single image super-resolution (HYSISR) is applied using multi resolution blending.
   - The Laplacian pyramid is applied to MISR and HYSISR and the Gaussian pyramid is applied to the mask.
   - The Blending ratio determines the influence of each input image in the output.
   - The output image’s pixels are in linear combination to the input pixels.
   - The composite image at level zero is to take pixels by adding synthetic high frequencies.

2) Segment the output hybrid super-resolution (HYSR) image.
   - Gaussian Kernel is used to recognize the sharpness of pixels.
   - Clustering take place to threshold the value score and the pixel values by DBSCAN.
   - Input the clusters created by DBSCAN,

3) Apply super-resolution reconstruction (SRR) technique to the segmented image.
Figure 4.6: Light Field Hybrid Super-resolution method. The method from the proposed approach in (Farag et al., 2017) where it shows a combination of two classical approaches HYISIR and MISR and merge them to gain a better resolution image. Image segmentation separates the objects in the background and foreground to enhance the sharpness of the merged image.

The diagram below shows an overall overview of the LFBMSR approach clarifies the approach of LFHYSR, which was explained in chapter three, and the LFBMSR, which has been implemented in this chapter.
Algorithm: Hybrid Approach for SR

**Input:** View Point images input $y^{(k)}$ extracted from the RAW image

1. To increase the VPs resolution by factor (4x4) do the following
   - a. Apply directional adapting interpolation
2. Extract the disparity information from the $y^{(k)}$
3. To apply LFBMSR, do the following:
   - a. Construct blocks $n_a \times n_a \times k \times k$
   - b. Search in the reference VPs, a number of similar blocks to $B$
   - c. Aggregation or mapping and 4D transform

1. To increase the VPs quality by factor (8x8), do the following:
   - a. For method one (MISR) $I_{MISR}$
   - b. For method two (HYSISR) $I_{HYSISR}$
2. To merge two images, do the following
   - a. Add pixel values of one layer with the other $I_{HYSR}$
3. To extract and segment the mask of object $I_{HYSR}$ do the following
   - a. Identify sharp pixels for each pixel in $I_{HYSR}$
   - b. Cluster generation from pixels
   - c. All clusters are linked to one contiguous region
   - d. Similar color to all neighbors creates set of regions $R_{color}$ in each pixel
   - e. For each region in $R_{color}$, relevance score is updated
   - f. Eliminate noise to generate segmented image
4. Apply SRR to the segmented image
4.4 Experiments and Simulation Results

The experiments for this chapter are applied to check the performance of the LFBMSR technique to prove that the resolution of all tested images is enhanced to the target resolution and all competitive methods that used LF images have been exceeded. To illustrate the proposed LFBMSR method’s performance, a set of experiments have been constructed. The Matlab-15a environment is utilized for all the experiments. The experiments use the real data (that is raw images), \( L(u, v, s, t) \). The ‘Toy’ and ‘Numbers’ images are obtained by the plenoptic camera (Hahne, 2016), whereas the Lego, Truck and Flower images are taken from Stanford University (Vaibhav, 2008). Consequently, several views from the raw image (i.e., 3x3, 4x3, 5x4 and 5x5) are used, as it is free to choose from the 9x9 extracted views; each view is 281 x 186. Subsequently, all VPs are interpolated to 1124x744 pixels using directional adaptive interpolation by factor 4x4. These HR images are generated as an input to the proposed LFBMSR method. Consequently, the image resolution is simplified, and the efficient sharpness of the interpolated HR images preserved. Respectively, disparity information is estimated using these HR images implementing the algorithm in (Sabter et al., 2015), then this information is used to facilitate BMSR by narrowing down the block search area in neighbouring VPs. Thus, the disparity pattern relies on the VPs of the scene from the plane of the LF. Furthermore, a LF block of dimension \( n_a \times n_a \times k \times k \) that corresponds to a 4D block was considered, as opposed to the BM3D, which searches and combines blocks only within a single image. Whereas the proposed light field block matching super-resolution (LFBMSR) approach operates across multiple views, exploiting similar blocks from different depths. Each VP is partitioned into 4D macro blocks, where the maximum displacement within the search space is \( \pm 7 \) pixels in both vertical and horizontal directions for all data used in Numbers, Toy,
Lego, Truck and Flower. Finally, to obtain the final super-resolved image by factor 8x8 equals to 2248 x 1488 pixels, the recent proposed LF hybrid super-resolution (LFHYSR) approach in (Farag et al., 2017) is used.

During implementation, multiple iterations were applied to choose an ideal search window size $FW$, (i.e., $\pm 7$, $\pm 11$ and $\pm 23$). Firstly, the pixels behaviour was tested to locate the best block match in all views, in order to determine if the proposed method performed better with or without exploiting the disparity information. Secondly, by using general BM, binary images were developed and exploited to identify the pixels, so as to discover which pixel chose the best match for each view as this was left again without constraint. Thirdly, this iteration was repeated for all competing methods, MISR, HYSISR and the method in (Wanner and Goldluecke, 2014).

While the depth was changing slightly, some of pixels were not visible because of occlusion, which can confuse block matching. Consequently, the BM method used without exploiting the depth was too random. Therefore, a better depth estimation method is utilised from (Sabter et al., 2015) as it performed better in terms of results, compared to the one without exploiting the disparity information (See Appendix B). Furthermore, after several iterations to find the best results, several window sizes $\pm 7$ and $\pm 11$ were selected in the LFBMSR approach and the ideal one was window size $\pm 7$. The outcome results have increased because of the method of depth estimation which allowed occlusion edges to be identified and treated occlusion explicitly. To calculate the efficiency of the proposed approach, a reconstructed reference image was created from the VPs of all data, such that each VP had $140 \times 93$ pixels image resolution. Furthermore, images were interpolated to 2248 x 1488 pixels in order to test all competitive methods, including the proposed one with the reference image.
Figure 4.7: Raw and View point Images (VPS). (Top) Raw image captured by the plenoptic camera (Hahne, 2016). (Bottom) A multi-spectral light field with 9 x 9 viewpoints (VPs) rearranged as a collection of several views.

The proposed approach is compared with the LFSR state-of-the-art methods and evaluated with regard to visual impression and objective quality. To measure the quality, PSNR, SSIM and computational complexity have been used. The method was compared to bicubic interpolation, MISR (Park et al., 2003) HYSISR (Xian et al., 2015) and (Wanner and Goldluecke, 2014).
4.5 Performance Evaluation

The performance evaluation outcomes of the proposed LFBMSR technique are presented in two sections. Firstly, section 4.5.1 presents the results of the directional adaptive interpolation method to enlarge the VPs by factor 4x4, furthermore the HR images were exploited to the proposed LFBMSR method. Secondly, the results in section 4.5.2 are compared based on the objective as well as visual quality. In order to gain a better performance compared to the recent methods, numerous results are illustrated after various iterations. Moreover, as highlighted in section 4.3, only 3x3, 4x3, 5x4 and 5x5 views were used because as the number of views increases the quality drops sharply, which is due to additional information added. Therefore, the implementation was stipulated to have a minimum of 3x3 and a maximum of 5x5 views. The visual results are presented in section 4.5.1 and 4.5.2, respectively. Note that the visual results are only for the ‘final estimation’ of the LFBMSR approach with the 3x3 and 5x5 views. Subsequently, the numeric comparison results are presented.

4.5.1 Directional Adaptive Interpolation Visual Results

This section shows the initial step applied to the proposed LFBMSR method, which enlarges the VPs resolution by factor 4x4. The size of each VP before the application of this approach was 281 x 186 pixels. Consequently, after applying the method, the VPs resolution has increased to 1124 x 744 pixels. The directional adaptive technique was compared with the bicubic interpolation and MISR (Park et al., 2003) to identify which of the methods have superior performance (For numerical results See Appendix B). In all the cases, the directional adaptive technique shows superior performance compared to other interpolation methods.
(bicubic interpolation method inclusive). Figures 4.8, 4.9 and 4.10 show the visual results of the central VPs of the ‘Numbers’ ‘Toy’ and ‘Lego’ images.

The central VP of the first tested LF image which is called ‘Numbers’ is shown in top image and a close up of the applied methods are shown below.

![The central view image ‘Numbers’](image)

**Figure 4.8:** Magnified *Numbers* image (a) Bicubic interpolation (b) (Kumar and Liyakathunisa, 2010) (c) (Velisavljevic, 2008).
Figure 4.9: Magnified Toy image (a) Bicubic interpolation (b) (Kumar and Liyakathunisa, 2010) (c) (Velisavljevic, 2008).
4.5.2 Proposed LFBMSR Visual and Numerical Results

The performance evaluation outcomes of the second step in the LFBMSR technique are presented in this section. The HR images shown in section 4.5.1 are exploited as an input to the proposed LFBMSR method. Furthermore, to enhance the resolution of the input HR images by factor 8x8 (2248 x 1488 pixels), the method implemented in chapter three is used. Consequently, the main contribution in this chapter is to use BM methods with SR to enhance LF images. Many iterations were applied to achieve the best images that could exceed other
existing methods. For instance, one of the iterations is to select the suitable VPs to be processed. In this study the method has been tested using a different number of views. For example, 3x3 views are used first and subsequently 4x3, 5x4 and 5x5. In Figures 4.11, 4.12, 4.13, 4.14 and 4.15 the visual results of 3x3 and 5x5 VPs are presented, which are the minimum and maximum number of views used to the proposed method. Subsequently, the numeric results shown in the Tables 4.1, 4.2, 4.3, 4.4 and 4.5.

The datasets used to test the proposed method are Numbers, Toy, Lego, ‘Truck’ and ‘Flowers’ which are selected to show the comparison results for the SR. Two datasets ‘Numbers’ and ‘Toy’ where captured by our plenoptic camera (Hahne, 2016) and the Lego, ‘Truck’ and ‘Flowers’ were taken by Stanford University (Vaibhav, 2008). As mentioned above, any datasets can be chosen to be tested by the proposed approach, but the selected datasets are higher and better quality compared to many plenoptic images and more specifically is that the most recent studies have used these datasets. The visual outcomes of the ‘Numbers’ image is shown in Figure 4.11. The implemented technique shows significant improvement compared to the other approaches, including (Wanner and Goldluecke, 2014). Furthermore, Figure 4.12 shows the outcomes of the image Toy. According to the visual comparison, the proposed approach outperforms other images results. Lastly, the comparison results of the Lego image shown in Figure 4.13, indicates that the performance of the proposed approach exceeds other competing methods. Moreover, the results of 5x5 views demonstrate that the image quality is better than 3x3 views in all the datasets used. Figure 4.14 and 4.15 shows the outcomes of the Truck and Flowers images. According to the visual comparison, the proposed approach outperforms other images results.
The ‘Numbers’ dataset in Figure 4.1 is captured by the plenoptic camera (Hahne, 2016). The quality is evaluated with other competing methods. This shows the results of 3x3 and 5x5 VPs.

**Figure 4.11**: The output images of all competitive methods and close ups. (a) and (b) 3x3 and 5x5 VPs of the HYSISR (Xian et al., 2015), (c) and (d) 3x3 and 5x5 VPs of MISR, (Park et al., 2003), (e) and (f) 3x3 and 5x5 VPs of (Wanner and Goldluecke., 2014), (g) and (h) 3x3 and 5x5 VPs of the proposed LFBMSR.
The ‘Toy’ dataset captured by the plenoptic camera (Hahne, 2016). The quality is evaluated with other competing methods. Figure 4.12 shows the ‘Toy’ image of 3x3 and 5x5 VPs.

<table>
<thead>
<tr>
<th>(3x3) VPs</th>
<th>(5x5) VPs</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
</tbody>
</table>

**Figure 4.12**: The output images of all competing methods and close ups. (a) and (b) 3x3 and 5x5 VPs of the HYSISR (Xian et al., 2015), (c) and (d) 3x3 and 5x5 VPs of MISR, (Park et al., 2003), (e) and (f) 3x3 and 5x5 VPs of (Wanner and Goldluecke., 2014), (g) and (h) 3x3 and 5x5 VPs of the proposed LFBMSR.
The ‘Lego’ image is captured by Stanford University (Vaibhav, 2008). The quality is evaluated with other competitive methods. Figure 4.13 shows the ‘Lego’ image of 3x3 and 5x5 VPs.

*Figure 4.13:* The output images of all competitive methods and close ups. (a) and (b) 3x3 and 5x5 VPs of the HYSISR (Xian et al., 2015), (c) and (d) 3x3 and 5x5 VPs of MISR (Park et al., 2003), (e) and (f) 3x3 and 5x5 VPs of (Wanner and Goldluecke., 2014), (g) and (h) 3x3 and 5x5 VPs of the proposed LFBMSR.
The ‘Truck’ image is captured by Stanford University (Vaibhav, 2008). The quality is evaluated with other competitive methods. Figure 4.13 shows the ‘Truck’ image of 3x3 and 5x5 VPs.

Figure 4.14: The output images of all competitive methods and close ups. (a) and (b) 3x3 and 5x5 VPs of the HYSISR (Xian et al., 2015), (c) and (d) 3x3 and 5x5 VPs of MISR (Park et al., 2003), (e) and (f) 3x3 and 5x5 VPs of (Wanner and Goldluecke, 2014), (g) and (h) 3x3 and 5x5 VPs of the proposed LFBMSR.
The ‘Flower’ image is captured by Stanford University (Vaibhav, 2008). The quality is evaluated with other competitive methods. Figure 4.13 shows the ‘Rose’ image of 3x3 and 5x5 VPs.

Figure 4.15: The output images of all competitive methods and close ups. (a) and (b) 3x3 and 5x5 VPs of the HYSISR (Xian et al., 2015), (c) and (d) 3x3 and 5x5 VPs of MISR (Park et al., 2003) (e) and (f) 3x3 and 5x5 VPs of (Wanner and Goldluecke., 2014), (g) and (h) 3x3 and 5x5 VPs of the proposed LFBMSR.
To summarise the visual results, the proposed method was evaluated against recent LFSR algorithms and compared with respect to objective quality, as well as visual impression. It was found to outperform all comparison state-of-the-art methods with a total average achieved over MISR, HYSISR and (Wanner and Goldluecke, 2014). This was achieved by searching for the best similar blocks in all VPs and super resolving the final image using the proposed LFHYSR approach (Farag et al., 2017). The images below illustrate the results of the final estimation images (5x5) views as it shows the best results compared to all views used.

**Figure 4.16**: (a) Shows central view point images extracted from raw images (Upper). Enhanced resolution images obtained after using the proposed LFBMSR method (Lower). (b) A close-up of all competitive super-resolved images.
After showing the evaluation of the visual quality comparison, the numerical results presented are based on the evaluation of PSNR and SSIM, as it is the most popular matrices to evaluate the methods of SR. Table 4.1 shows the PSNR values of LFBMSR with a 7x7 window size. In each number of views used, there are basic and final estimation images. The basic estimation is the first iteration applied for the BM step, whereas, the final estimation step is done based on the first estimation step as it considered the second iteration of the BM step. The final estimation results perform better in all cases compared to the basic estimation. On the other hand, the results of the proposed approach are highlighted in Table 4.1, Table 4.2, Table 4.3, Table 4.4 and Table 4.5 respectively, which indicates better performance compared to other methods.

For instance, in Table 4.1 the ‘Numbers’ image shows that the highlighted PSNR and SSIM results exceed all other existing methods in all number of views utilized. Furthermore, the LFBMSR final estimation 5x5 views outperform the result in (Wanner and Goldluecke, 2014) by 0.49 dB. Similarly, the ‘Toy’ image outcomes exceed the result in (Wanner and Goldluecke, 2014) by 0.44 dB. Also, the ‘Lego’ data set outperform the result in (Wanner and Goldluecke, 2014) by 0.44 dB. Furthermore, in Table 4.4 the ‘Truck’ image, shows that the highlighted PSNR and SSIM results exceeds all other existing methods in all number of views utilized. The LFBMSR final estimation 5x5 views outperform the result in (Wanner and Goldluecke, 2014) by 0.51 dB. For the last dataset, the ‘Flower’ image shown in Table 4.5, show that the highlighted PSNR and SSIM results exceed all other existing methods in all number of views utilized. The LFBMSR final estimation 5x5 views outperform the result in (Wanner and Goldluecke, 2014) by 0.45 dB. Finally, all results prove that the LFBMSR technique gained superior performance comparable to existing, competing methods.
Table 4.1: Displays the values of PSNR dB and SSIM compared among the LFSR approaches

<table>
<thead>
<tr>
<th>Numbers Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFBMSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>3x3</td>
<td>34.24</td>
<td>0.9320</td>
<td>35.55</td>
<td>0.9367</td>
<td>35.68</td>
</tr>
<tr>
<td>Final estimation</td>
<td>3x3</td>
<td>34.30</td>
<td>0.9329</td>
<td>35.60</td>
<td>0.9376</td>
<td>35.72</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>4x3</td>
<td>34.10</td>
<td>0.9316</td>
<td>35.46</td>
<td>0.9357</td>
<td>35.57</td>
</tr>
<tr>
<td>Final estimation</td>
<td>4x3</td>
<td>34.15</td>
<td>0.9321</td>
<td>35.50</td>
<td>0.9364</td>
<td>35.63</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>5x4</td>
<td>34.34</td>
<td>0.9335</td>
<td>35.53</td>
<td>0.9363</td>
<td>35.68</td>
</tr>
<tr>
<td>Final estimation</td>
<td>5x4</td>
<td>34.49</td>
<td>0.9341</td>
<td>35.64</td>
<td>0.9371</td>
<td>35.75</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>5x5</td>
<td>34.65</td>
<td>0.9344</td>
<td>35.68</td>
<td>0.9378</td>
<td>35.82</td>
</tr>
<tr>
<td>Final estimation</td>
<td>5x5</td>
<td>34.71</td>
<td>0.9354</td>
<td>35.73</td>
<td>0.9388</td>
<td>35.96</td>
</tr>
<tr>
<td>Our Previous LFHYSR presented in Chapter 3</td>
<td>bicubic</td>
<td>33.64</td>
<td>0.9341</td>
<td>35.44</td>
<td>0.9370</td>
<td>35.65</td>
</tr>
</tbody>
</table>

The next table shows another data set used for the experiment, based on the ‘Toy’ image and is also captured by the plenoptic camera (Hahne, 2016). Similarly, the outcome of the comparison shows that the proposed approach outperforms all existing methods, including LFHYSR, MISR (Park et al., 2003), HYSISR (Xian et al., 2015) and (Wanner and Goldluecke, 2014) using PSNR and SSIM.
Table 4.2: Displays the values of PSNR dB and SSIM compared among the LFSR approaches.

<table>
<thead>
<tr>
<th>Toy Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFBMSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
</tr>
<tr>
<td>Basic estimation 3x3</td>
<td>34.10</td>
<td>0.9368</td>
<td>35.20</td>
<td>0.9376</td>
<td>35.24</td>
<td>0.9386</td>
</tr>
<tr>
<td>Final estimation 3x3</td>
<td>34.18</td>
<td>0.9374</td>
<td>35.24</td>
<td>0.9387</td>
<td>35.29</td>
<td>0.9399</td>
</tr>
<tr>
<td>Basic estimation 4x3</td>
<td>33.91</td>
<td>0.9357</td>
<td>35.09</td>
<td>0.9371</td>
<td>35.22</td>
<td>0.9383</td>
</tr>
<tr>
<td>Final estimation 4x3</td>
<td>33.95</td>
<td>0.9364</td>
<td>35.16</td>
<td>0.9375</td>
<td>35.24</td>
<td>0.9387</td>
</tr>
<tr>
<td>Basic estimation 5x4</td>
<td>34.14</td>
<td>0.9383</td>
<td>35.27</td>
<td>0.9389</td>
<td>35.45</td>
<td>0.9396</td>
</tr>
<tr>
<td>Final estimation 5x4</td>
<td>34.18</td>
<td>0.9391</td>
<td>35.43</td>
<td>0.9397</td>
<td>35.57</td>
<td>0.9411</td>
</tr>
<tr>
<td>Basic estimation 5x5</td>
<td>34.21</td>
<td>0.9398</td>
<td>35.42</td>
<td>0.9411</td>
<td>35.54</td>
<td>0.9421</td>
</tr>
<tr>
<td>Final estimation 5x5</td>
<td>34.27</td>
<td>0.9411</td>
<td>35.53</td>
<td>0.9417</td>
<td>35.70</td>
<td>0.9429</td>
</tr>
</tbody>
</table>

Our Previous LFHYSR presented in Chapter 3 5x5 | bicubic | HYSISR | MISR | (Wanner and Goldluecke) | LFHYSR |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>33.27</td>
<td>0.9148</td>
<td>35.27</td>
<td>0.9398</td>
<td>35.48</td>
</tr>
</tbody>
</table>

The next table shows the PSNR dB and SSIM values compared with other LFSR approaches. The next table shows another data set used for the experiment, which is the ‘Lego’ image, as captured at Stanford University (Vaibhav, 2008). Similarly, the outcome of the comparison shows that the proposed approach outperforms all existing methods, including LFHYSR, MISR (Park et al., 2003), HYSISR (Xian et al., 2015) and (Wanner and Goldluecke, 2014).
Table 4.3: Displays the values of PSNR dB and SSIM compared among the LFSR approaches.

<table>
<thead>
<tr>
<th>Lego Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFHYSR) Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR SSIM</td>
<td>PSNR SSIM</td>
<td>PSNR SSIM</td>
<td>PSNR SSIM</td>
<td>PSNR SSIM</td>
<td>PSNR SSIM</td>
</tr>
<tr>
<td>Basic</td>
<td>3x3</td>
<td>34.32 0.9388</td>
<td>35.42 0.9396</td>
<td>35.44 0.9406</td>
<td>36.04 0.9414</td>
<td>36.37 0.9434</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>3x3</td>
<td>34.40 0.9394</td>
<td>35.46 0.9408</td>
<td>35.49 0.9417</td>
<td>36.14 0.9431</td>
<td>36.41 0.9442</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>4x3</td>
<td>34.23 0.9378</td>
<td>35.32 0.9392</td>
<td>35.44 0.9403</td>
<td>35.88 0.9411</td>
<td>36.31 0.9417</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>4x3</td>
<td>34.17 0.9386</td>
<td>35.38 0.9398</td>
<td>35.47 0.9408</td>
<td>35.98 0.9414</td>
<td>36.34 0.9429</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>5x4</td>
<td>34.35 0.9403</td>
<td>35.48 0.9310</td>
<td>35.68 0.9414</td>
<td>36.32 0.9431</td>
<td>36.67 0.9445</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>5x4</td>
<td>34.40 0.9411</td>
<td>35.65 0.9417</td>
<td>35.76 0.9420</td>
<td>36.41 0.9438</td>
<td>36.72 0.9452</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>5x5</td>
<td>34.43 0.9417</td>
<td>35.63 0.9393</td>
<td>35.72 0.9439</td>
<td>36.37 0.9444</td>
<td>36.80 0.9468</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>5x5</td>
<td>34.90 0.9367</td>
<td>35.86 0.9396</td>
<td>36.21 0.9423</td>
<td>36.67 0.9456</td>
<td>37.11 0.9532</td>
</tr>
<tr>
<td>estimation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Our Previous LFHYSR</td>
<td>5x5</td>
<td>bicubic</td>
<td>HYSISR</td>
<td>MISR (Wanner and Goldluecke)</td>
<td>LFHYSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33.71 0.9350</td>
<td>35.53 0.9381</td>
<td>35.71 0.9411</td>
<td>36.21 0.9433</td>
<td>36.82 0.9510</td>
</tr>
</tbody>
</table>

The next table shows the PSNR dB and SSIM values compared with other LFSR approaches.

The next table shows another data set used for the experiment, which is the ‘Truck’ image, as captured at Stanford University (Vaibhav, 2008). Similarly, the outcome of the comparison shows that the proposed approach outperforms all existing methods, including LFHYSR, MISR (Park et al. 2003), HYSISR (Xian et al., 2015) and (Wanner and Goldluecke, 2014).
The next table shows the PSNR dB and SSIM values compared with other LFSR approaches.

The next table shows another data set used for the experiment, which is the ‘Flower’ image, as captured at Stanford University (Vaibhav, 2008). Similarly, the outcome of the comparison shows that the proposed approach outperforms all existing methods, including LFHYSR, MISR (Park et al. 2003), HYSISR (Xian et al., 2015) and (Wanner and Golduecke, 2014).
Table 4.5: Displays the values of PSNR dB and SSIM compared among the LFSR approaches

<table>
<thead>
<tr>
<th>Flower Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFBMSR)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PSNR</td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
<td>PSNR</td>
<td>SSIM</td>
</tr>
<tr>
<td>Basic estimation 3x3</td>
<td>34.38</td>
<td>0.9394</td>
<td>35.51</td>
<td>0.9403</td>
<td>35.53</td>
<td>0.9411</td>
<td>36.13</td>
</tr>
<tr>
<td>Final estimation 3x3</td>
<td>34.49</td>
<td>0.9398</td>
<td>35.52</td>
<td>0.9411</td>
<td>35.56</td>
<td>0.9422</td>
<td>36.24</td>
</tr>
<tr>
<td>Basic estimation 4x3</td>
<td>34.29</td>
<td>0.9384</td>
<td>35.40</td>
<td>0.9404</td>
<td>35.52</td>
<td>0.9408</td>
<td>35.96</td>
</tr>
<tr>
<td>Final estimation 4x3</td>
<td>34.27</td>
<td>0.9391</td>
<td>35.45</td>
<td>0.9406</td>
<td>35.55</td>
<td>0.9413</td>
<td>36.08</td>
</tr>
<tr>
<td>Basic estimation 5x4</td>
<td>34.44</td>
<td>0.9407</td>
<td>35.57</td>
<td>0.9314</td>
<td>35.70</td>
<td>0.9419</td>
<td>36.35</td>
</tr>
<tr>
<td>Final estimation 5x4</td>
<td>34.49</td>
<td>0.9416</td>
<td>35.74</td>
<td>0.9425</td>
<td>35.84</td>
<td>0.9426</td>
<td>36.53</td>
</tr>
<tr>
<td>Basic estimation 5x5</td>
<td>34.51</td>
<td>0.9424</td>
<td>35.71</td>
<td>0.9405</td>
<td>35.81</td>
<td>0.9442</td>
<td>36.45</td>
</tr>
<tr>
<td>Final estimation 5x5</td>
<td>34.98</td>
<td>0.9456</td>
<td>35.95</td>
<td>0.9409</td>
<td>36.30</td>
<td>0.9427</td>
<td>36.74</td>
</tr>
<tr>
<td>Our Previous LFHYSR</td>
<td>5x5</td>
<td>bicubic</td>
<td>HYSISR</td>
<td>MISR</td>
<td>(Wanner and Goldluecke)</td>
<td>LFHYSR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33.71</td>
<td>0.9350</td>
<td>35.53</td>
<td>0.9381</td>
<td>35.71</td>
<td>0.9411</td>
</tr>
</tbody>
</table>
4.6 Summary

This chapter has implemented a novel LFBMSR method which combines several existing super and interpolation methods. At first, the directional adaptive interpolation is applied to the VPs to increase the image size by factor 4x4. Different approaches have been used to increase the VPs resolution. However, simulation tests have been done to compare the quality between the directional adaptive methods and other interpolation methods, such as the bicubic interpolation method. As a result, the directional adaptive method outperformed the competitive methods. Next, the HR images of the VPs are used as an input. Afterward, the extracted disparity information from the VPs is exploited to find the correct distance between the VPs. This estimation method has been changed after a number of iterations to get the best results, as the method used previously was not ideal. After that, the iteration of BM over VPs is employed to search for the matches blocks with the reference image. Initial iteration starts from a chosen VP, represented as the reference VP. Then, one VP is selected from the LF image as the central. At this point, the search window is created to search in the neighboring VPs for the best blocks that match the reference block in the central VP image chosen. The size of the window $n_a \times n_a$ is an angular search window to run all VPs together. The iteration is then done under reference VP of the 2D blocks. Next, the combination of the 2D blocks result in forming 4D blocks, which are created by the reference VP and the neighboring VPs. After that, the rest of the remaining pixels in neighboring VPs need to be processed for a second iteration. After fully processing the neighboring VPs with current reference, consideration is given to it the remaining VPs to generate a new reference VP.

After LFBMSR is complete, the proposed LFHYSR implemented in chapter three is used to conclude the process, which combines the BM (HYSISR) and BM (MISR) classical approaches.
The combination of MISR and HYSISR images are then stored in the mask object using the blending addition mode, where the output image’s pixels are in linear combination to the input pixels. The composite image at level zero is to take pixels by adding synthetic high frequencies; after adding pixels, it must result with a HR image HYSR. Image segmentation is then applied to the HYSR image, as to segment the super-resolved image plays an important role in defining an object from its background and is commonly used to partition an image into meaningful regions and enhance the edges. Next, SRR is utilised to the last segmented image, to get the final LFHYSR image, which is consider the last step of this method.

To conclude this chapter, a well-structured, SR method, applied to LF images has been implemented. Furthermore, it has provided a solution to the limitations of the LR images of the plenoptic camera. Moreover, the literature review has presented the recent studies applied to BM, as well as the depth map extraction methods. Also, it has explained the proposed approach and illustrated the simulation of the implemented method. Consequently, the results were evaluated against recent LFSR algorithms and compared with respect to objective quality, as well as visual impression. The LFBMSR final estimation 5x5 views outperform all the comparison state-of-the-art methods, with a total average achieved over MISR, with respect to PSNR, in tune of up to 1.06 dB and 0.97 dB; as well as 0.0108 and 0.0029 for SSIM, respectively, for the ‘Numbers’ and ‘Toy’ images, and 0.90 dB , 1.10 dB 0.89 dB; and 0.0109, 0.0096, 0.0108 for the ‘Lego’ ‘Truck’ and ‘Flower’ image. This is achieved by searching for the best similar blocks in all VPs and super resolving the final image using the recent proposed LFHYSR approach. The following chapter present different benchmark multi-focused images and the proposed method from this chapter is applied to these refocused images to enhance the resolution.
5. Super-resolution of All-in-Focus Light Field Images

The LF camera captures image in both spatial and angular resolution. It enables capabilities that are new, including camera aperture adjustment, the view point images (VPs) shift, an estimation that is in-depth and post-capture refocusing (Ng et al., 2005). In addition, the plenoptic camera provides a feature of extracting refocused images from the light field (LF) raw image. All these refocused images are collected to create an all-in-focus image. Consequently, the main idea is to apply the proposed technique in chapter four to super resolve the all-in-focus image. The proposed technique’s performance is evaluated objectively and numerically using the two popular testing methods PSNR and SSIM with the recent proposed approaches in (Park et al., 2003) and (Boominathan et al., 2015). This work presents the last contribution developed in this thesis.
5.1 Introduction

Using the capture of directional and spatial light ray distribution in space, new capabilities have been enabled by LF cameras. These capabilities include the refocused images which generated by integrated existing developed methods. As a result of the efficiency of LF camera transmission light, the microlens array-based LF camera is considered the most preferred design (Boominathan et al., 2015). However, the microlens array is based on the LF cameras as they suffer from spatial LR which blocks the adoption of LF cameras extensive as well the microlens array shares one sensor when capturing both angular and spatial information (Farrugia et al., 2017).

Subsequently, all of the refocused images extracted from the camera suffer from low resolution (LR), thus, improving the spatial resolution is necessary. The contribution of this chapter is different from that of chapters four because the datasets used in this chapter are refocused images whereas in chapter four the images used were the view point images (VPs) that extracted from plenoptic camera. But the method applied in chapter four is also used for this chapter to super-resolve the quality of the multi refocused images. This chapter has been structured accordingly, in section 5.2 the limitations of the recent LF refocusing SR methods applied to plenoptic images to enhance the resolution is explained. Section 5.3 illustrates the extraction of the refocused images obtained from the raw images. After that, in section 5.4 the proposed technique is shown in detail. Next section presents the results gained from the simulation experiments. Last section evaluates the performance of the applied method.
5.2 Limitations of Plenoptic Camera in Refocusing

The authors (Isaksen et al., 2000) demonstrated virtual refocusing from the LF. In a recent study, the original LF rendering graphics file is usually referred to as synthetic aperture photograph (Levoy and Hanrahan, 1996). It is suggested that it can take into consideration the interesting variation of the disposed focal plane. For example, the VP surface toward the main lens. However, these refocus features have two main problems within the interior design of the camera. First, large camera arrays are not suitable for capturing the necessary LF dataset and can be used for long-time scans associated with traditional photography or moving cameras for spontaneous shooting. Second, because of inadequate inspection of the virtual focal point gap such as the camera hole, the outcomes tend to indicate high complexity in the obscured area. As compared to the conventional camera, LF camera is careful calibration. Moreover, aliasing is significantly reduced with the optical design by the integration of the light that is passed through the aperture (Ng et al., 2005).

Theoretically, it was one of the first attempt to formulate the refocus accurately based on the actual image processing performed in the camera (Ng et al., 2005). Their attempt results were closer to the traditional images than those in the previous works. To this day, this relationship is more abstract (virtual lens sometimes more than 1 meter in diameter). The main goal is to qualitatively replicate the limited DoF digital effects.

The cost-effective of the LF images is provided by cameras uses microlens array based on LF. Microlens arrays have two main limitations, the first one is baseline that is very narrow as well as a spatial resolution that is very low. Low spatial resolution ensures that it has limited general-purpose use as well as the applicability of microlens array based. The second limitation is the accuracy and the range of depth estimation, which is limited by narrow
baseline (Boominathan et al., 2015). Consequently, a combined system known as a “hybrid stereo imaging” designed by the authors (Alam and Gunturk, 2017; Farrugia et al., 2017), the system has a composition of a LF and conventional camera. Moreover, limitations associated with narrow baseline and spatial resolutions in microlens array based on LF cameras are addressed by this hybrid system, as well as the capabilities of LF imaging are preserved. On the low spatial resolution, several methods addressed this issue in microlens array.

5.3 Super-resolution Approaches Applied to All-in-Focus Images

Examples of the recent approaches include the technique of the application of SR restoration applied to VP images and the suggestions of Bayesian framework as well as GMM and learning based (Farrugia et al., 2017; Bishop et al., 2012). The authors included dictionary learning and natural networks that are deep convolutional. There has also been the utilization of techniques known as Fourier domain as well as wave optics that are based 3D dimension deconvolution methods.

All systems of LF imaging that have sensors are considered limited due to the spatial-angular resolution. Consequently, there are performance limitations in the restoration methods mentioned above, and this in addition to the costs of the computation.

Another approach that can be used to improve spatial resolution is based on the use of the two-camera system that is considered as a hybrid by using a HR camera and a LF camera. The system allows in merging of the images as a way of improving spatial resolution (Farrugia et al., 2017). However, the limitation of this technique is based on dictionary learning in which HR image obtained from a conventional camera, is stored as HR dictionary patch. Consequently, another technique of down-sampling these HR patches, is to extract the
features that have LR, and this leads to forming patch dictionary of LR. During the reconstruction of SR, obtaining the HR of LF takes place, by enhancing the LR patches in the image, which is done through the determination of HR in the dictionary. The decomposition of the images that have HR and complex steerable filters of pyramids are used. Further, the up-sampling of the depth map in the LF and bilateral up-sampling that is jointly is used.

The ability to estimate the depth is considered as essential aspect of the LF cameras’ features. However, microlens -attached to the LF cameras have no depth accuracy due to the narrow baseline. There has been a study of the relationship between the narrow baseline and the accuracy of estimation of the depth in a stereo system (Zhao et al., 2015; Zhao et al., 2018).

**Figure 5.1:** Left, Raw LF. Right, View point images that are decoded (Alam and Guntrunk, 2017).

Several methods based on software are implemented recently to super resolve the limitation the LF images, as well as HR patches, using GMM, dictionary learning and texture priors. There can be an application to these methods which combines the conventional and LF cameras (Mukati and Gunturk, 2018), and in this case, there is a formation of an image on microlens array using the objective lens. To form an image on the sensor, the microlens as well as the objective lens cooperate. In standard plenoptic cameras, the angular information is captured
by the microlens pixels, and there is a formation of camera’s spatial resolution from a number of micro-lens. For the case of focused LF camera, there is a formation of a perspective image on the sensor from microlenses. Consequently, (Mukati and Gunturk, 2018) have proposed methods that are designed with an aim of improving the refocused LF images spatial resolution. Consequently, (Gul and Gunturk, 2018) stated that systems of hybrid including cameras that are conventional and LF cameras have also been proved to enhance the LF spatial resolutions. Cameras end up having different VPs when a conventional camera and LF camera have a different optical axis, and this results in regions that are occluded. In order to prevent the issue of occlusion, there is a use of beam splitters in front of cameras as this ensures that cameras have the same optical axis.

In comparison to the above approaches, a micro scanning technique is proposed by (Gul and Gunturk, 2018). There is a fusion of multiple input images as a way of improving spatial resolution, and this happens in the image restoration of the SR that is a multi-frame. Proper among the input images are all that is required, and this is in order to have the locations of pixel diversity sampling. Rather than taking many images and having a movement that is proper among these images for covering the space of pixel sampling, there can be a movement induction directly on the sensor, and at the same time, the recording of samples that are necessary can take place. For instance, the moving of an image sensor can take place, moving horizontally, vertically and diagonally, this producing four different images. An image with x4 of the original resolution is formed after compositing these images together. The idea of micro-scanning has been used in many fields such as digital photography and video recording. One disadvantage of this technique is that there is a need for the scene to be static.
when multiple images are being captured so that registration processes can be avoided. There can be alleviation to the issue of the dynamic scene to few points while there is a use of piezoelectric actuators and synchronized with image sensors that are very fast. An object that moves in the scene can also be captured when there is an application of the technique that can incorporate the step of registration into the process of restoration.

In contrast to the methods applied in applications of LFSR discussed above, the proposed technique is applied to refocused images using the advantage of the technique used in chapter four. According to the literature of this study, the existing techniques have improved the hardware part of the plenoptic camera rather than apply SR techniques to enhance the resolution of the LF images, like (Bishop et al., 2009; Bishop et al., 2012; Boominathan et al., 2015). The authors enhanced image details by designing a Variational Bayesian framework, by increasing the resolution of the measured LF, which exploits prior knowledge about the scene and estimates both the LF and HR depth map which extracts additional information from available data. Their method provides image formation by distinguishing the plenoptic camera point spread function (PSF) under Gaussian optics assumptions for a several-depth scene.

However, in the proposed method, extracting the multi refocused images is applied first from the raw image as shown in Figure 5.2 in the next section. Then the proposed method from chapter four is applied to these refocused images to enhance the resolution.

5.4 Extracting Focused Images from the raw Data

In (Adelson and Wang, 1991; Dansereau, 2014; Ng et al., 2005), the display can extract VPs from the plenoptic camera by gathering all pixels having the same small-scale image location.
Based on the given symbol, the one-dimensional VP $E_i [S_j]$ having the index $i$ is given by the following equation.

$$E_i [s_j] = Ef_S [s_j, u_c + i]$$  \hspace{1cm} (5.1)

Ever since $i$ is considered as sufficient indicator of a VP image of a one-dimensional line, $u$ and $c$ are omitted by the suffix $E_i$. Eq (1.5) demonstrates that the successful determination is equivalent to the quantity of miniaturized scale focal points. Figure 5.2 displays the outcome of generating the two-dimensional VP image $E_{(i,g)}$ from the space and directional region index variable $[S_i, t_h]$ and $[u_{c+j}, v_{c+g}]$. Enhancement pixels indicate that samples within a particular micro image correspond to corresponding viewpoint positions within the camera array.

Figure 5.2: Extracting VPs which leads to focused images. (a) Calibrate the original image obtained by plenoptic camera. By $[u_{c+j}, v_{c+g}]$ the micro image sample passes through the $[S_i, t_h]$ and pixel indices in the micro image. Here, $M = 3$. (b) The extracted two-dimensional sub opening image $E(i, g)$. Every colour represents different VP. Coordinates $[u_{c+j}, v_{c+g}]$ indexes the perspective image and $[S_i, t_h]$ indexes associated spatial pixels.
Ever since the capture of the original plenoptic camera doesn’t naturally characterize $E_{fs} [S_j, u_{c+i}]$ indexes representation, the LF image is considered to have two conventional sensor dimensions $[x, y]$. Index transformation is as follows:

$$K = j \times M + c + i$$

(5.2)

In the horizontal dimension formed by $[x]$

$$[x] = [x_j \times M + c + i] = E_{fs} [S_j, u_{c+i}]$$

(5.3)

Similarly, vertical index transformation is also possible.

$$y = h \times M + c + g$$

(5.4)

So;

$$[y] = [yh \times M + c + g] = [t_h, v_{c+g}]$$

(5.5)

Create a 4D LF symbol $[Sj, Uc + i, th , Vc + g]$ that are applied to the conventional two-dimensional sampling formula $[x, y]$. Counts from index $O$.

Isaksen’s revelation enables to carefully change of the focal plane of an image by applying SR and merging the pixels of a refocused images by procuring the LF. Therefore, since the previously extracted VPs describes the refocusing process from the VP in Figure 5.2, it can be used for refocusing. The one-dimensional refocusing process from the VP mathematically becomes as follows:
\[ E'[S_j] = \sum_{i=-c}^{c} E_i[S_{j-a(c+i)}], \quad a \in \mathbb{Q} \] (5.6)

This indicates that the refocusing from VP also contains the sum of the directional samples.

Eq. 5.6 requires performing VP extraction beforehand when the original image captured from the plenoptic camera refocusing. This results in two separate images rendering forms. For instance, computerized centring of the condition changes. As appeared, since the sub opening extraction is just a reordering of the pixel esteems, the process can be executed during the integration process in order to adapt this process to such a situation in a single process.

**Figure 5.3:** Refocusing scheme, method based on VP images (Hahne, 2016).

In order to achieve refocusing from the multi-view image, the presented VPs are superimposed on each other and appear slightly intertwined while the superimposed pixels are coordinated. The move parameter influences the refocusing separation. The sub-opening image is featured with a shading coordinating the LF model (Hahne, 2016). Figure 5.4 presents one of the plenoptic raw image.
The close-up view displays micro images. The distance between the refocused images is validated by (Hahne, 2016) as their method compete the recent studies according to their outcomes. In the proposed approach the advantages of these distances have been used. For instance, Figure 5.5 shows the refocused images $E_{a''}[S_j, t_h]$ at various slices. As explained in the above section how to extract multi refocused images using the technique of (Hahne, 2016). Because it is the first phase in the method. The next figure shows the Spiderman dataset in different refocused images.

**Figure 5.5:** From the 9 refocused images extracted these are 1/9 and 9/9 images. The furthest and the closest.
5.5 Overview of the Proposed Refocussing SR method

Input LF image

Extract VPs → VP selection process → Up sampling shift and integration

Check if $i = \text{noShift}$

Image $i$ generated at different depth

Apply LFBMSR by selecting window size on image $i$

Check if $i = \text{noShift}$

Output different depth of LF images

Extract the super-resolved window and store the window in new arrays and record the $i$ value with highest resolution

Disparity information is calculated from $i = \text{shift}$

All-in-focused image is generated with stored SR windows

Output All-in-focused image

Output different depth of LF images

Figure 5.6: Flowchart of the proposed method.
The steps applied to the proposed technique are in the flowchart shown in Figure 5.6. Initial stage is to obtain the multi refocused images from the reconstructed LF image, which suffer from LR. Second stage is applied to these refocused images so that their resolution increases by 4x4 using the adaptive interpolation method in (Velisavljevic, 2008). The following step is to apply LFBMSR proposed in chapter four which is capable of enhancing spatial resolution of super-resolved LF images while preserving low computational complexity, the disparity information is estimated from the HR images using the algorithm in (Sabter et al., 2015). This information is then used to facilitate block matching SR by narrowing down the block search area in neighbouring VPs. Thus, the disparity pattern relies on the VPs of the scene from the plane of the LF. Further, to obtain the final super-resolved image, the LFHYSR approach that illustrated in chapter three, is used. In order to get a SR all-in-focused images, extracting of super-resolved window is done as well as store the window in new arrays to record the value with highest resolution images.

5.6 Experiments and Results

The differences between the experiments carried out in this chapter and chapter four, is that the method in this chapter was applied to the refocused images to achieve a SR all-in-focused images. However, chapter four experimental results applied the method to enhance the resolution of the VPs. The proposed LF refocusing super resolution method has exceeded all competitive methods applied to LF images.

A set of experiments have been made. The Matlab-15a environment is utilized for all the experiments. The experiments use the real data (i.e., raw images), \( L(u, v, s, t) \). The Spider and Numbers images are obtained by the standard plenoptic camera (Hahne, 2016), whereas Lego
image is taken from Stanford University (Vaibhav, 2008). The nine refocused images are reconstructed from the tested images, each refocused image is 281 x 186 pixels as shown above in Figure 5.5.

The experiments are carried out into six phases:

1) The adaptive interpolation method is used to increase all of the refocused images to 4x4 factors as well as to simplify the image resolution and to preserve efficient sharpness in the interpolated high resolution (HR) images. Theses HR images are generated as an input to the proposed LFBMSR method illustrated in chapter four.

2) Respectively, disparity information is estimated (Sabter et al., 2015), using these HR images using. Then this information is used to facilitate block matching SR by narrowing down the block search area in neighbouring VPs. Thus, the disparity pattern relies on the VPs of the scene from the plane of the LF.

3) LF block of dimension $n_a \times n_a \times k \times k$ that corresponds to a 4D block was considered as opposed to the BM3D, which searches and combines blocks only within a single image, whereas the proposed LFBMSR approach operates across multiple views, exploiting similar blocks from different depths.

4) Each VP is partitioned into 4D macro blocks, where the maximum displacement within the search space is ±7 pixels in both vertical and horizontal directions for all data used Spiderman, Lego and Numbers.

5) To obtain the final super-resolved image by factor 8x8 equals to 2248 x 1488 pixels, the recent proposed LF hybrid super-resolution (LFHYSR) approach in (Farag et al., 2017) is used.
6) To test the quality of the images, a reference image is generated using an interpolation method to increase the image size to the target enhanced images of the \((Spiderman, Lego\) and \(Numbers\)).

5.7 Performance Evaluation

The performance is evaluated using a reference image, which generated using linear interpolation method to enhance the resolution to match the target competitive super-resolved images in all tested images \((Spiderman, Lego\) and \(Numbers\)). The PSNR measurement matrices technique is used to compare the results between existing approaches and the proposed approach. The other approaches are implemented by (Park et al., 2003) and (Boominathan et al., 2015). SSIM is another technique to evaluate the results and used to clarify the best performance in all cases. Over all, according to the PSNR and SSIM results the proposed approach exceeds the other two techniques based on subjectively and objectively. The outcomes of the implemented refocusing SR method are presented in two sections. The visual results are compared in section 5.6.1. The numerous results are illustrated after various iterations to gain the best performance comparable to the recent methods. As mentioned in section 5.4.1, only \(1/9, 3/9, 7/9\) and \(9/9\), refocused images were used, as if more views were used, the quality is dropping sharply and that’s due additional information unneeded. Therefore, it was settled with a min of five refocused images.
5.7.1 Visual Results

The following images that show at the top of the following pages have different focus. For instance, 1/9, 3/9, 5/9, 7/9, 9/9. The images resolution is 281 x 186. The proposed SR method is applied to these five multi refocused images in order to enhance the resolution 8x8 factor. The datasets (Spiderman and Numbers) are LF images captured by the standard plenoptic camera in (Hahne, 2016). Whereas, the LF (Lego) image is obtained from Stanford University (Vaibhav, 2008).

The four images at the bottom of the next page below, are the super resolved ‘Spiderman’ all-in-focused image. The quality is evaluated with other competitive methods. The proposed method outperforms the existing approach as can be seen in Figure 5.7. The two methods show poorness in quality compared to the proposed. Second data used is the ‘Lego’ image. The visual quality of the proposed technique exceeds the two comparison methods used as shown in Figure 5.8. Third image shown is the ‘Numbers’ image which is evaluated with other competitive methods as presented in Figure 5.9. Overall, in all cases the proposed technique has achieved and outperformed the two approaches by (Park et al., 2003) and (Boominathan et al., 2015).
Figure 5.7: All-in-focus super-resolved images. Show the proposed method compares to the existing approaches. (a) Reference, (b) Park et al., 2003), (c) (Boominathan et al., 2015) (d) Proposed. The close-up images show a magnified image to clear the details of the images.
Figure 5.8: All-in-focus super-resolved images. Show the proposed method compares to the existing approaches. (a) Reference, (b) (Park et al., 2003) (c) (Boominathan et al., 2015), (d) Proposed. The close-up images show a magnified image to clear the details of the images.
**Numbers**

Figure 5.9: All-in-focused super-resolved refocused images showing the proposed and the compared images. (a) Reference, (b) (Park et al., 2003) (c) (Boominathan et al., 2015) (d) Proposed.
5.7.2 Numerical Results

Table 5.1. Total average of PSNR and SSIM results.

<table>
<thead>
<tr>
<th>Methods</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spider man</td>
<td>Lego</td>
</tr>
<tr>
<td>(Park et al., 2003)</td>
<td>35.51</td>
<td>35.95</td>
</tr>
<tr>
<td>(Boominathan et al., 2015)</td>
<td>35.85</td>
<td>36.10</td>
</tr>
<tr>
<td>Proposed</td>
<td>36.21</td>
<td>36.49</td>
</tr>
<tr>
<td>Average gain over Boominathan</td>
<td>0.36</td>
<td>0.39</td>
</tr>
</tbody>
</table>

5.8 Summary

To conclude this chapter, a well-structured method for LF images has been accomplished and addressed a solution of the problem of the LR outputs of those refocused images. Furthermore, the chapter presented the recent studies applied to refocused images and the work applied in order to enhance the superiority of the image’s resolution. Moreover, it has explained the proposed approach and clarified the simulation of the implemented method as well as the outcome results were calculated and compared to the existing methods applied to all-in-focus images and evaluated in terms of visual impression as well as objective quality and exceed all state-of-the-art approaches with a total average gained over (Boominathan et al., 2015) with respect to PSNR in tune of up to 0.36 dB, 0.39 and 0.34 dB; in addition to 0.0016, 0.0028 and 0.0019 for SSIM, correspondingly, for the ‘Spiderman’ ‘Lego’ and ‘Numbers’ datasets.
6. Conclusion

Advances in technology of the plenoptic camera have gained important attention from both the research community and industry, so as to provide efficient scene refocusing at different depths, without additional computational complexity. This type of camera is based on the concept initially introduced and further developed in existing studies, as highlighted in the literature review of this work. One of the common ways to represent LF data is to consider it a matrix of VPs, where each VP captures a 2D slice of the LF. In contrast to classical cameras, plenoptic cameras are capable of capturing images (2D LF slices) at multiple VPs at the same time, owing to a system of multi-array lenses; a significant leap forward in camera hardware technology.
Recently, the available plenoptic camera technology is providing limited depth-of-field (DoF) and low spatial image resolution due to its efficiency of transmission light. For example, the final super resolution (SR) of the Lytro camera is restricted to only 300 x 300 pixels. Nevertheless, such cameras are limited due to additional hardware restrictions and often, its resolution is not achievable in practice. Additionally, the design known as the microlens array-based LF camera, which is considered the most preferred design, suffers from spatial LR because it shares one sensor when capturing angular and spatial information, a situation that raises the need for post-processing techniques, such as SR.

6.1 Research Significance and Contributions Summary

The significance of the proposed research is to develop SR and interpolation techniques to solve the limitation of the LR images extracted by the plenoptic cameras. Moreover, applying the proposed methods for the LF images allow a better and higher quality to all LF images captured by plenoptic cameras. It also allows applications in areas such as the film industry to process images at a higher resolution.

Several approaches have been proposed in the literature to enhance this limited spatial resolution. For instance, epipolar plane images (EPIs) have been applied to measure variational methods and disparity maps in order to calculate a LF image. However, measuring the variational and disparity maps using VPs is more accurate. For instance, other studies exploited the VP information for the LF denoising application by extending the block matching 3D (BM3D) filter (Dabov et al., 2007; Lebrun, 2012). In contrast to recent studies, the method proposed in this work exploits BM to light-field super-resolution (LFSR), rather than motion estimation in classical video processing, and is the first to do so.
Consequently, a novel approach termed light field hybrid super-resolution (LFHYSR) which provides a high-quality reconstruction framework has been proposed. The method uses free (that is, open) LR VPs, which are a combination of two classical SR techniques for efficient application to plenoptic images. Further, a segmentation technique is used on the output of a hybrid super-resolution (HYSR) image into the objects of interest. Afterward, the sparse representation technique has been applied to super resolve the segmented image. This technique helped to improve the quality by decreasing computations for LF images and extracting significant features from the objects of interest.

A further development of LFHYSR, called the “light field block matching super resolution” (LFBMSR) technique, is the proposal that it would be capable of enhancing the spatial resolution of super-resolved LF images, while preserving low computational complexity. The approach consisted of two steps: firstly, the image interpolation technique is applied to each sub-aperture image to interpolate the spatial resolution by 4x4. Secondly, the disparity information is estimated from the HR images using the algorithm in (Park et al., 2003). This information is then used to facilitate BM SR by narrowing down the block search area in neighbouring VPs. Thus, the disparity pattern relies on the VPs of the scene from the plane of the LF. Further, to obtain the final super-resolved image, the proposed LFHYSR is used.

Lastly, different input of datasets called ‘refocused images’, which also suffer from LR, are extracted from LF images of the plenoptic camera and have been enhanced by the LFHYSR and LFBMSR techniques. Further, the output of the proposed LF refocusing method is capable of presenting a high-quality super resolution image that is all-in-focused.

In conclusion, tests and evaluations have been conducted to ensure quality comparison between the proposed techniques, as well as techniques of the existing methods. The
proposed schemes were evaluated using PSNR and SSIM. The results have demonstrated remarkable performance compared to the existing methods.

6.2 Future Work

Consequent to the implementation of the novel methods in this work, improvements that could further reduce the computational complexity of the LF images are suggested. These improvements can be categorised into the design of the plenoptic camera and the enhancement of the LF images spatial resolution.

Part of the future work is to improve the performance to the implemented method for LFBMSR (specifically, taken at Stanford University (Vaibhav, 2008) and the University of Bedfordshire (Hahne, 2016) and for factors that have high amplification, the suggestion is to amend the proposed improvement issue so as to join visions from multi-views SR. Another hypothesis can be suggested, which proposes in the same angular position that a LR VP can take the advantages of the neighbouring VPs to produce a HR VP. This could affect the BM step of the proposed method, as the LR residual error up sampling would also come from the surrounding VPs, not only the corresponding ones.

Furthermore, object extraction (also known as object segmentation) could be enhanced, if VPs are created with the SR method. Likewise, enhanced BM, for example, LFBMSR could be actualised for the purpose of developing a quality disparity map, which could help to minimize the faults caused to the ultimate object segmentation. Certain disparity map information can be extracted effectively with the desired object by developing a region of interest method implementation.
The plenoptic camera could be improved if an extension of the hardware design would be investigated, through the effect of the dynamic-real time-of imaging and produce important information to suit varies applications in the field of engineering. Another suggested approach is to apply SR to Lytro camera images with zoom feature because it has not been investigated, consequently directing analyses by considering the zoom feature could show insight to unexplored regions of the Lytro imaging with regards to designing applications of engineering. Finally, another part of the hypothesis is to dispose of the standard SR strategy that is combined with the camera, for instance, hybrid hardware and software techniques. Consequently, testing the application of the coefficient’s sensor attached with various SR methods, which are considered commercial and non-commercial, to see the outcomes effects. Subsequently, to get the effect on the output image with several SR approaches, another SR approach, particularly for LF images must be improved.
References


Bishop, T. and Favaro, P. (2009) “Plenoptic depth estimation from multiple aliased views”. In Computer Vision Workshops (ICCV Workshops), IEEE 12th International Conference on, pp. 1622–1629


174


175


Georgiev and Lumsdaine, (2009) “Depth of Field in Plenoptic Cameras”, Adobe Systems, Inc. San Jose, CA USA 2 Indiana University, Bloomington, IN USA


Yan, W. Ren, Y. Guo, R. Wang, and X. Cao. (2017) “Image de-blurring via extreme channels prior”. In CVPR.


Yu, L. Xu, Y. Xu, H. Yang, x. (2013) “Self-Example Based Super-Resolution with Fractal-Based Gradient Enhancement” Department of Electronic Engineering, Shanghai Jiao tong University, Shanghai 200240, China


A. Appendix

Quality Validation of Datasets and Mask Generation using Laplacian Pyramid

To support the results in chapter three, the test was applied to explore which method perform better around the edge region and which perform better around the smooth region of the MISR method as well as HYSISR. Thus, to select the better one as the target image for the combination to be accomplished. The results show that the MISR is better around the smooth region and the edge relation in all cases. Therefore, the MISR image is used as the target image. The output combined HYSR is also been tested to see if it performs better than the MISR ands HYSISR. Overall, the result of HYSR image out perform the PSNR results of competing methods based on smooth and edge regions.
1) Results below are visual and numerical of 'Toy' image (HYSISR and MISR methods)

<table>
<thead>
<tr>
<th></th>
<th>HYSISR</th>
<th></th>
<th>MISR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Laplacian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 Results of **HYSISR** image (Edge region) (Smooth region)

<table>
<thead>
<tr>
<th>Toy Image</th>
<th>T=0.080</th>
<th>T=0.090</th>
<th>T=0.10</th>
<th>T=0.15</th>
<th>T=0.19</th>
<th>T=0.29</th>
<th>T=0.39</th>
<th>T=0.49</th>
<th>T=0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edge region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>33.26</td>
<td>33.28</td>
<td>33.29</td>
<td>33.40</td>
<td>33.42</td>
<td>33.44</td>
<td>33.48</td>
<td>33.50</td>
<td>33.52</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9631</td>
<td>0.9632</td>
<td>0.9635</td>
<td>0.9638</td>
<td>0.9635</td>
<td>0.9634</td>
<td>0.9632</td>
<td>0.9630</td>
<td>0.9629</td>
</tr>
<tr>
<td><strong>Smooth region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>32.35</td>
<td>32.36</td>
<td>32.37</td>
<td>32.39</td>
<td>32.40</td>
<td>32.44</td>
<td>32.46</td>
<td>32.45</td>
<td>32.43</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9596</td>
<td>0.9597</td>
<td>0.9599</td>
<td>0.9602</td>
<td>0.9598</td>
<td>0.9599</td>
<td>0.9598</td>
<td>0.9597</td>
<td>0.9596</td>
</tr>
</tbody>
</table>

Table 2 Results of **MISR** image (Edge region) (Smooth region)

<table>
<thead>
<tr>
<th>Toy Image</th>
<th>T=0.080</th>
<th>T=0.090</th>
<th>T=0.10</th>
<th>T=0.15</th>
<th>T=0.19</th>
<th>T=0.29</th>
<th>T=0.39</th>
<th>T=0.49</th>
<th>T=0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Edge region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>34.10</td>
<td>33.85</td>
<td>33.87</td>
<td>33.98</td>
<td>33.99</td>
<td>34.02</td>
<td>34.08</td>
<td>34.09</td>
<td>34.12</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9677</td>
<td>0.9674</td>
<td>0.9678</td>
<td>0.9681</td>
<td>0.9682</td>
<td>0.9680</td>
<td>0.9679</td>
<td>0.9677</td>
<td>0.9676</td>
</tr>
<tr>
<td><strong>Smooth region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>32.94</td>
<td>32.86</td>
<td>32.87</td>
<td>32.89</td>
<td>32.90</td>
<td>32.94</td>
<td>32.93</td>
<td>32.94</td>
<td>32.96</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9639</td>
<td>0.9640</td>
<td>0.9639</td>
<td>0.9641</td>
<td>0.9641</td>
<td>0.9643</td>
<td>0.9640</td>
<td>0.9639</td>
<td>0.9641</td>
</tr>
</tbody>
</table>
2) Results below are visual and numerical of ‘Numbers’ image (HYSISR and MISR methods)
### Table 3 Results of HYSISR image (Edge region) (Smooth region)

<table>
<thead>
<tr>
<th>Numbers</th>
<th>Image</th>
<th>T=0.080</th>
<th>T=0.090</th>
<th>T=0.10</th>
<th>T=0.15</th>
<th>T=0.19</th>
<th>T=0.29</th>
<th>T=0.39</th>
<th>T=0.49</th>
<th>T=0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge region</td>
<td>PSNR</td>
<td>33.59</td>
<td>33.62</td>
<td>33.69</td>
<td>33.98</td>
<td>33.14</td>
<td>33.41</td>
<td>33.48</td>
<td>33.52</td>
<td>33.52</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9700</td>
<td>0.9700</td>
<td>0.9701</td>
<td>0.9705</td>
<td>0.9707</td>
<td>0.9709</td>
<td>0.9708</td>
<td>0.9707</td>
<td>0.9704</td>
</tr>
<tr>
<td>Smooth region</td>
<td>PSNR</td>
<td>32.15</td>
<td>32.16</td>
<td>32.15</td>
<td>32.11</td>
<td>32.10</td>
<td>32.08</td>
<td>32.09</td>
<td>32.10</td>
<td>32.11</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9638</td>
<td>0.9639</td>
<td>0.9640</td>
<td>0.9637</td>
<td>0.9638</td>
<td>0.9636</td>
<td>0.9637</td>
<td>0.9637</td>
<td>0.9637</td>
</tr>
</tbody>
</table>

### Table 4 Results of MISR image (Edge region) (Smooth region)

<table>
<thead>
<tr>
<th>Numbers</th>
<th>Image</th>
<th>T=0.080</th>
<th>T=0.090</th>
<th>T=0.10</th>
<th>T=0.15</th>
<th>T=0.19</th>
<th>T=0.29</th>
<th>T=0.39</th>
<th>T=0.49</th>
<th>T=0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge region</td>
<td>PSNR</td>
<td>34.89</td>
<td>34.93</td>
<td>34.99</td>
<td>34.28</td>
<td>34.44</td>
<td>34.69</td>
<td>34.74</td>
<td>34.79</td>
<td>34.80</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9717</td>
<td>0.9718</td>
<td>0.9719</td>
<td>0.9723</td>
<td>0.9725</td>
<td>0.9726</td>
<td>0.9725</td>
<td>0.9725</td>
<td>0.9721</td>
</tr>
<tr>
<td>Smooth region</td>
<td>PSNR</td>
<td>33.39</td>
<td>33.38</td>
<td>33.37</td>
<td>33.34</td>
<td>33.31</td>
<td>33.31</td>
<td>33.33</td>
<td>33.34</td>
<td>33.36</td>
</tr>
<tr>
<td></td>
<td>SSIM</td>
<td>0.9662</td>
<td>0.9663</td>
<td>0.9664</td>
<td>0.9658</td>
<td>0.9657</td>
<td>0.9657</td>
<td>0.9655</td>
<td>0.9654</td>
<td>0.9657</td>
</tr>
</tbody>
</table>
3) Results below are visual and numerical for ‘Numbers and Toy’ image. (HYSR method)
### Table 5 Results of HYSR Image (Edge region) (Smooth region)

<table>
<thead>
<tr>
<th>HYSR (Numbers) Image</th>
<th>T=0.080</th>
<th>T=0.090</th>
<th>T=0.10</th>
<th>T=0.15</th>
<th>T=0.19</th>
<th>T=0.29</th>
<th>T=0.39</th>
<th>T=0.49</th>
<th>T=0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>35.34</td>
<td>35.437</td>
<td>35.39</td>
<td>35.44</td>
<td>35.51</td>
<td>35.63</td>
<td>35.72</td>
<td>35.81</td>
<td>35.87</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9861</td>
<td>0.9865</td>
<td>0.9868</td>
<td>0.9862</td>
<td>0.9861</td>
<td>0.9863</td>
<td>0.9868</td>
<td>0.9865</td>
<td>0.8662</td>
</tr>
<tr>
<td>Smooth region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>34.40</td>
<td>34.39</td>
<td>34.38</td>
<td>34.36</td>
<td>34.34</td>
<td>34.31</td>
<td>34.35</td>
<td>33.36</td>
<td>34.38</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9850</td>
<td>0.9851</td>
<td>0.9852</td>
<td>0.9855</td>
<td>0.9856</td>
<td>0.9860</td>
<td>0.9862</td>
<td>0.9858</td>
<td>0.9856</td>
</tr>
</tbody>
</table>

### Table 6 Results of HYSR Image (Edge region) (Smooth region)

<table>
<thead>
<tr>
<th>HYSR (Toy) Image</th>
<th>T=0.080</th>
<th>T=0.090</th>
<th>T=0.10</th>
<th>T=0.15</th>
<th>T=0.19</th>
<th>T=0.29</th>
<th>T=0.39</th>
<th>T=0.49</th>
<th>T=0.59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edge region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>34.67</td>
<td>34.69</td>
<td>34.75</td>
<td>34.78</td>
<td>34.84</td>
<td>34.88</td>
<td>34.90</td>
<td>34.92</td>
<td>34.97</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9771</td>
<td>0.9775</td>
<td>0.9778</td>
<td>0.9782</td>
<td>0.9781</td>
<td>0.9783</td>
<td>0.9778</td>
<td>0.9675</td>
<td>0.9672</td>
</tr>
<tr>
<td>Smooth region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>33.39</td>
<td>33.38</td>
<td>33.37</td>
<td>33.34</td>
<td>33.31</td>
<td>33.30</td>
<td>33.32</td>
<td>33.34</td>
<td>33.36</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9752</td>
<td>0.9753</td>
<td>0.9754</td>
<td>0.9762</td>
<td>0.9761</td>
<td>0.9760</td>
<td>0.9759</td>
<td>0.9758</td>
<td>0.9761</td>
</tr>
</tbody>
</table>
B. Appendix

Adaptive interpolation method and Depth Estimation Technique

In this work, several tests were carried out. The tests are divided into two main points to support the novelty of the proposed LFBMSR technique from chapter four.

1) **Apply adaptive interpolation method**: To the sub-aperture images and measure the quality of other existing interpolation methods using PSNR and SSIM. According to the results, the interpolation approach of (Velisavljevic, 2008) performed the best (See results below).

2) **Depth estimation method**: Is needed to extract the disparity information from the sub-aperture images used in chapter four. After several iterations to find the best results, several window sizes were selected in the LFBMSR approach and the ideal one is window size 7. (See results below)
Subsequently, enlarge the sub-aperture images is needed to increase the LR images to get HR images. Therefore, to use these HR images in the proposed method as an input rather using LR image. The decision of choosing the method to be used, is first to compare between the performance of some interpolation existing methods to clear out the performance of the methods. Thus, I have selected two meaningful methods to increase the resolution x4. This was applied to all of the data sets that I have tested, the Numbers, Toy and Lego images. The chosen methods are from (Velisavljevic, 2008) and (Kumar and Liyakathunisa, 2010).

(1) **Apply adaptive interpolation method**

Consequently, the visual and numerical results are presented below. The method in (Velisavljevic, 2008), gained better resolution in terms of PSNR and visual results. Therefore, this method is utilized in the proposed method to increase the resolution before SR is applied.
• The central view image (Numbers) (280 x 170)

1. After bicubic interpolation applied, (1124 x 744) (x4)
2. Kumar [80], (1124 × 744) (x4)

3. Velisavljevic [144], (1124 × 744) (x4)
• The central view image (Toy) (280 x 170)

1. After bicubic interpolation applied, (1124 x 744) (x4)
2. Kumar [80], (1124 × 744) (x4)

3. Velisavljevic [144], (1124 × 744) (x4)
1. **Bicubic Interpolation**

2. Kumar [80], (1124 × 744) (x4)

3. Veličković [144], (1124 × 744) (x4)
1) Bicubic Interpolation

2. Kumar [80], (1124 x 744) (x4)

3. Velisavljevic [144], (1124 x 744) (x4)
The method gained better resolution in terms of visual results is the method of (Velisavljevic, 2008). Therefore, this method is utilized in the proposed LFBMSR to increase the resolution.

All tables below show the PSNR and SSIM results after of both approaches applied to the sub-aperture images. The proposed LFBMSR is applied and the disparity information has been exploited after changing the method of depth map estimation.

Table 1. (Kumar and Liyakathunisa, 2010) (x4) (window size 7) (Numbers and Toy data sets)

(PSNR)

<table>
<thead>
<tr>
<th>Numbers Image (x4)</th>
<th>Number of Views</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFHYSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>35.55</td>
<td>35.68</td>
<td>36.11</td>
<td>36.34</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3)</td>
<td>35.60</td>
<td>35.72</td>
<td>36.17</td>
<td>36.40</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>35.68</td>
<td>35.82</td>
<td>36.20</td>
<td>36.81</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5)</td>
<td>35.73</td>
<td>35.96</td>
<td>36.44</td>
<td>36.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Toy Image (x4)</th>
<th>Number of Views</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFHYSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>35.20</td>
<td>35.24</td>
<td>35.84</td>
<td>36.19</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3)</td>
<td>35.24</td>
<td>35.29</td>
<td>35.96</td>
<td>36.24</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>35.42</td>
<td>35.54</td>
<td>36.18</td>
<td>36.61</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5)</td>
<td>35.53</td>
<td>35.70</td>
<td>36.29</td>
<td>36.73</td>
</tr>
</tbody>
</table>
Table 2. (Kumar and Liyakathunisa, 2010) (x4) (window size 7) (Numbers and Toy data sets) (SSIM)

<table>
<thead>
<tr>
<th>Numbers Image (x4)</th>
<th>Number of Views</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFHYSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>0.9367</td>
<td>0.9383</td>
<td>0.9413</td>
</tr>
<tr>
<td></td>
<td>Final estimation</td>
<td>(3x3)</td>
<td>0.9376</td>
<td>0.9392</td>
<td>0.9423</td>
</tr>
<tr>
<td></td>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>0.9378</td>
<td>0.9406</td>
<td>0.9431</td>
</tr>
<tr>
<td></td>
<td>Final estimation</td>
<td>(5x5)</td>
<td>0.9388</td>
<td>0.9413</td>
<td>0.9438</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Toy Image (x4)</th>
<th>Number of Views</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFHYSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>0.9376</td>
<td>0.9386</td>
<td>0.9396</td>
</tr>
<tr>
<td></td>
<td>Final estimation</td>
<td>(3x3)</td>
<td>0.9387</td>
<td>0.9399</td>
<td>0.9412</td>
</tr>
<tr>
<td></td>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>0.9411</td>
<td>0.9421</td>
<td>0.9425</td>
</tr>
<tr>
<td></td>
<td>Final estimation</td>
<td>(5x5)</td>
<td>0.9417</td>
<td>0.9429</td>
<td>0.9434</td>
</tr>
</tbody>
</table>
### Table 3. (Velisavljevic, 2008) (x4) (window size 7) (Numbers and Toy data sets) (PSNR)

<table>
<thead>
<tr>
<th>Numbers Image (x4)</th>
<th>Number of Views</th>
<th>Basic estimation (3x3)</th>
<th>Final estimation (3x3)</th>
<th>Basic estimation (5x5)</th>
<th>Final estimation (5x5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>PSNR</strong></td>
<td><strong>PSNR</strong></td>
<td><strong>PSNR</strong></td>
<td><strong>PSNR</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BMSR (HYSISR)</td>
<td>BMSR (MISR)</td>
<td>BMSR (Wanner and Goldluecke)</td>
<td>BMSR (LFHYSR)</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>36.55</td>
<td>36.68</td>
<td>37.11</td>
<td>37.34</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3)</td>
<td>36.60</td>
<td>36.72</td>
<td>37.17</td>
<td>37.40</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>36.68</td>
<td>36.82</td>
<td>37.20</td>
<td>37.81</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5)</td>
<td>36.73</td>
<td>36.96</td>
<td>37.44</td>
<td>37.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Toy Image (x4)</th>
<th>Number of Views</th>
<th>Basic estimation (3x3)</th>
<th>Final estimation (3x3)</th>
<th>Basic estimation (5x5)</th>
<th>Final estimation (5x5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>PSNR</strong></td>
<td><strong>PSNR</strong></td>
<td><strong>PSNR</strong></td>
<td><strong>PSNR</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>BMSR (HYSISR)</td>
<td>BMSR (MISR)</td>
<td>BMSR (Wanner and Goldluecke)</td>
<td>BMSR (LFHYSR)</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>36.20</td>
<td>36.24</td>
<td>36.84</td>
<td>37.19</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3)</td>
<td>36.24</td>
<td>36.29</td>
<td>36.96</td>
<td>37.24</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>36.42</td>
<td>36.54</td>
<td>37.18</td>
<td>37.61</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5)</td>
<td>36.53</td>
<td>36.70</td>
<td>37.29</td>
<td>37.73</td>
</tr>
<tr>
<td>Numbers Image (x4)</td>
<td>Number of Views</td>
<td>BMSR (HYSISR)</td>
<td>BMSR (MISR)</td>
<td>BMSR (Wanner and Goldluecke)</td>
<td>BMSR (LFHYSR)</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>0.9467</td>
<td>0.9483</td>
<td>0.9513</td>
<td>0.9587</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3)</td>
<td>0.9476</td>
<td>0.9492</td>
<td>0.9523</td>
<td>0.9599</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>0.9478</td>
<td>0.9506</td>
<td>0.9531</td>
<td>0.9613</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5)</td>
<td>0.9488</td>
<td>0.9513</td>
<td>0.9538</td>
<td>0.9621</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Toy Image (x4)</th>
<th>Number of Views</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
<th>BMSR (Wanner and Goldluecke)</th>
<th>BMSR (LFHYSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic estimation</td>
<td>(3x3)</td>
<td>0.9476</td>
<td>0.9486</td>
<td>0.9496</td>
<td>0.9514</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3)</td>
<td>0.9487</td>
<td>0.9499</td>
<td>0.9512</td>
<td>0.9521</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5)</td>
<td>0.9511</td>
<td>0.9521</td>
<td>0.9525</td>
<td>0.9547</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5)</td>
<td>0.9517</td>
<td>0.9529</td>
<td>0.9534</td>
<td>0.9558</td>
</tr>
</tbody>
</table>
(2) Depth estimation method to select best window size for block matching

This step is applied to do more iterations to enhance the performance of block matching (BM) exploiting disparity information.

First attempt is to change the window size of the search window in all test images. Below are the different sizes window sizes selected to

Second attempt is to increase the block size of each window. Properly choosing them is for improve of crucial importance in order to have good quality results, the most important ones are:

- Size of Block
- Size of the search range
- Kind of the research

The test is divided to two sections:

(a) Results after disparity information are exploited

(b) The results below don’t exploit disparity information
(a) Results after disparity information is exploited

**NUMBERS IMAGE (WINDOW SIZE ±11 )**

- BMSR (Bicubic)
- BMSR (HYSISR)
- BMSR (MISR)

**PSNR VS NUMBER OF VIEWS**

**TOY IMAGE (WINDOW SIZE ±11)**

- BMSR (bicubic)
- BMSR (HYSISR)
- BMSR (MISR)

**PSNR VS NUMBER OF VIEWS**
The results below exploit disparity information (window size ±11)

<table>
<thead>
<tr>
<th>Numbers Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic estimation</td>
<td>(3x3) PSNR</td>
<td>33.15</td>
<td>34.49</td>
<td>34.54</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3) PSNR</td>
<td>33.29</td>
<td>34.62</td>
<td>34.71</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(4x3) PSNR</td>
<td>33.04</td>
<td>34.39</td>
<td>34.38</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(4x3) PSNR</td>
<td>33.14</td>
<td>34.54</td>
<td>34.59</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x4) PSNR</td>
<td>33.32</td>
<td>34.57</td>
<td>34.65</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x4) PSNR</td>
<td>33.46</td>
<td>34.69</td>
<td>34.76</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5) PSNR</td>
<td>33.51</td>
<td>34.79</td>
<td>34.83</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5) PSNR</td>
<td>33.55</td>
<td>34.89</td>
<td>35.12</td>
</tr>
<tr>
<td>Toy Image</td>
<td>Number of Views</td>
<td>BMSR (bicubic)</td>
<td>BMSR (HYSISR)</td>
<td>BMSR (MISR)</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(3x3) PSNR</td>
<td>33.09</td>
<td>34.11</td>
<td>34.17</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3) PSNR</td>
<td>33.16</td>
<td>34.24</td>
<td>34.39</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(4x3) PSNR</td>
<td>32.91</td>
<td>33.81</td>
<td>33.90</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(4x3) PSNR</td>
<td>32.97</td>
<td>33.88</td>
<td>33.97</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x4) PSNR</td>
<td>33.11</td>
<td>34.35</td>
<td>34.41</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x4) PSNR</td>
<td>33.17</td>
<td>34.44</td>
<td>34.56</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5) PSNR</td>
<td>33.15</td>
<td>34.41</td>
<td>34.48</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5) PSNR</td>
<td>33.21</td>
<td>34.50</td>
<td>34.69</td>
</tr>
</tbody>
</table>
The results below exploit disparity information (window size ±7)

<table>
<thead>
<tr>
<th>Numbers Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic estimation</td>
<td>(3x3) PSNR</td>
<td><strong>33.45</strong></td>
<td><strong>34.79</strong></td>
<td><strong>34.84</strong></td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3) PSNR</td>
<td><strong>33.69</strong></td>
<td><strong>34.85</strong></td>
<td><strong>34.91</strong></td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(4x3) PSNR</td>
<td><strong>33.34</strong></td>
<td><strong>34.51</strong></td>
<td><strong>34.64</strong></td>
</tr>
<tr>
<td>Final estimation</td>
<td>(4x3) PSNR</td>
<td><strong>33.59</strong></td>
<td><strong>34.69</strong></td>
<td><strong>34.73</strong></td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x4) PSNR</td>
<td><strong>33.62</strong></td>
<td><strong>35.21</strong></td>
<td><strong>35.34</strong></td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x4) PSNR</td>
<td><strong>33.68</strong></td>
<td><strong>35.33</strong></td>
<td><strong>35.46</strong></td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5) PSNR</td>
<td><strong>33.86</strong></td>
<td><strong>35.49</strong></td>
<td><strong>35.70</strong></td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5) PSNR</td>
<td><strong>33.97</strong></td>
<td><strong>35.62</strong></td>
<td><strong>35.78</strong></td>
</tr>
<tr>
<td>Toy Image</td>
<td>Number of Views</td>
<td>BMSR (bicubic)</td>
<td>BMSR (HYSISR)</td>
<td>BMSR (MISR)</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Basic estimation (3x3) PSNR</td>
<td>33.07</td>
<td>34.12</td>
<td>34.19</td>
<td></td>
</tr>
<tr>
<td>Final estimation (3x3) PSNR</td>
<td>33.10</td>
<td>34.23</td>
<td>34.31</td>
<td></td>
</tr>
<tr>
<td>Basic estimation (4x3) PSNR</td>
<td>32.93</td>
<td>34.21</td>
<td>34.23</td>
<td></td>
</tr>
<tr>
<td>Final estimation (4x3) PSNR</td>
<td>33.11</td>
<td>34.28</td>
<td>34.37</td>
<td></td>
</tr>
<tr>
<td>Basic estimation (5x4) PSNR</td>
<td>33.32</td>
<td>34.65</td>
<td>34.81</td>
<td></td>
</tr>
<tr>
<td>Final estimation (5x4) PSNR</td>
<td>33.57</td>
<td>34.84</td>
<td>34.96</td>
<td></td>
</tr>
<tr>
<td>Basic estimation (5x5) PSNR</td>
<td>33.41</td>
<td>35.35</td>
<td>35.41</td>
<td></td>
</tr>
<tr>
<td>Final estimation (5x5) PSNR</td>
<td>33.64</td>
<td>35.46</td>
<td>35.58</td>
<td></td>
</tr>
</tbody>
</table>
(b) The results below don’t exploit disparity information (window size ±7)

<table>
<thead>
<tr>
<th>Numbers Image</th>
<th>Number of Views</th>
<th>BMSR (bicubic)</th>
<th>BMSR (HYSISR)</th>
<th>BMSR (MISR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic estimation</td>
<td>(3x3) PSNR</td>
<td>33.34</td>
<td>35.19</td>
<td>35.24</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3) PSNR</td>
<td>33.41</td>
<td>35.29</td>
<td>35.51</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(4x3) PSNR</td>
<td>33.24</td>
<td>34.49</td>
<td>34.64</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(4x3) PSNR</td>
<td>33.31</td>
<td>34.71</td>
<td>34.76</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x4) PSNR</td>
<td>33.47</td>
<td>34.68</td>
<td>34.75</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x4) PSNR</td>
<td>33.63</td>
<td>34.79</td>
<td>34.81</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5) PSNR</td>
<td>33.73</td>
<td>35.51</td>
<td>35.58</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5) PSNR</td>
<td>33.85</td>
<td>35.58</td>
<td>35.64</td>
</tr>
<tr>
<td>Toy Image</td>
<td>Number of Views</td>
<td>BMSR (bicubic)</td>
<td>BMSR (HYISR)</td>
<td>BMSR (MISR)</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(3x3) PSNR</td>
<td>33.15</td>
<td>34.74</td>
<td>34.89</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(3x3) PSNR</td>
<td>33.23</td>
<td>34.91</td>
<td>34.93</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(4x3) PSNR</td>
<td>33.10</td>
<td>34.46</td>
<td>34.54</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(4x3) PSNR</td>
<td>33.13</td>
<td>34.51</td>
<td>34.61</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x4) PSNR</td>
<td>33.41</td>
<td>34.65</td>
<td>34.71</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x4) PSNR</td>
<td>33.53</td>
<td>34.73</td>
<td>34.77</td>
</tr>
<tr>
<td>Basic estimation</td>
<td>(5x5) PSNR</td>
<td>33.59</td>
<td>34.71</td>
<td>34.83</td>
</tr>
<tr>
<td>Final estimation</td>
<td>(5x5) PSNR</td>
<td>33.51</td>
<td>34.80</td>
<td>34.91</td>
</tr>
</tbody>
</table>