# IMAGE ENHANCEMENT THROUGH NEW TECHNIQUES IN COMPUTATIONAL PHOTOGRAPHY

A DISSERTATION SUBMITTED TO

THE GRADUATE SCHOOL OF

ENGINEERING AND NATURAL SCIENCES

OF ISTANBUL MEDIPOL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ELECTRICAL, ELECTRONICS ENGINEERING AND CYBER SYSTEMS

By

Muhammad Zeshan Alam

March, 2019

Image enhancement through new techniques in computational photography

By Muhammad Zeshan Alam

March, 2019

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

Prof. Dr. Bahadır Kürşat Güntürk(Advisor)

Prof. Dr. Hasan Fehmi Ateş

Assoc. Prof. Dr. Taner Eskil

Assist. Prof. Dr. Muhammed Fatih Toy

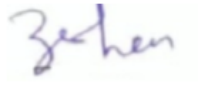
Assist. Prof. Dr. Muharrem Mercimek

Approved by the Graduate School of Engineering and Natural Sciences:

Prof. Dr. Talip Alp
Director of the Graduate School of Engineering and Natural Sciences

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

Name, Last Name: MUHAMMAD ZESHAN ALAM

Signature : 

# ABSTRACT

## IMAGE ENHANCEMENT THROUGH NEW TECHNIQUES IN COMPUTATIONAL PHOTOGRAPHY

Muhammad Zeshan Alam
Ph.D. in Electrical, Electronics Engineering and Cyber Systems
Advisor: Prof. Dr. Bahadır Kürşat Güntürk
March, 2019

Quality of a digital image depends on several factors operative during the image formation process, e.g. sensor defects, sensor dynamic range, poor spatial resolution, lens distortion, camera shake and object motion. This study focused on developing new techniques in computational photography for minimizing some degradations in digital images, including blurring, limited depth of field, low dynamic range, and insufficient resolution.

A flexible framework is developed for space-variant deblurring using a single degraded image. Coarse PSF estimation of image patches and PSF clustering are performed to identify regions of uniform blur in an image followed by PSF refinement, deconvolution, and fusion. Focus stacking and high dynamic range (HDR) imaging are combined to generate all-in-focus HDR image using multiple exposure and multiple focus images, captured through a camera array. The limited resolution problem is addressed in the context of light field imaging in two different ways: hybrid stereo imaging involving a regular camera and a light field camera and deconvolution based high-resolution light field extraction from a single image capture.

All the developed algorithms are tested on real datasets and both qualitative and quantitative comparisons have been made with the state-of-the art methods to show the superiority of the proposed algorithms.

*Keywords:* Space-variant image deblurring, light field imaging, computational photography, image enhancement, high dynamic range imaging, focus stacking.

# ÖZET

# HESAPLAMALI GÖRÜNTÜLEMEDE YENİ TEKNİKLERLE GÖRÜNTÜ İYİLEŞTİRME

Muhammad Zeshan Alam
Elektrik-Elektronik Mühendisliği ve Siber Sistemler, Doktora
Tez Danışmanı: Prof. Dr. Bahadır Kürşat Güntürk
Mart, 2019

Dijital bir görüntünün kalitesi, görüntü kaydetme sürecinde etkin olan sensör hataları, sensör dinamik aralığı, düşük uzamsal çözünürlük, merceklerden kaynaklanan bozulmalar, kamera sarsıntısı ve nesne hareketi gibi çeşitli faktörlere bağlıdır. Bu çalışma; dijital görüntülerde bulanıklık, sınırlı odak derinliği, düşük dinamik aralık ve yetersiz çözünürlük için bazı bozulmaları en aza indirmek için yeni teknikler geliştirmeye odaklanmaktadır.

Tek bir bozulmuş görüntüden uzamsal-değişken netleştirme amaçlı esnek bir algoritma geliştirilmiştir. Algoritma, görüntüdeki temel bulanıklık bölgelerini bulmaya yönelik kaba bir nokta dağılım fonksiyonu (PSF) kestirimi ve kümeleme, PSF kestirimini iyileştirme, dekonvolüsyon ve füzyon aşamalarının içerir. Odak istifleme ve yüksek dinamik aralıklı (HDR) görüntülemede kamera dizisiyle çekilen çoklu pozlama ve odaklama görüntüleri kullanılarak geniş odaklı ve yüksek dinamik aralıklı görüntü oluşturulmaktadır. Işık alanı görüntülemedeki sınırlı uzamsal çözünürlük sorunu iki farklı şekilde ele alınmıştır: normal bir kamera ve ışık alanı kamerası içeren hibrit görüntüleme ve dekonvolüsyon ile tek bir görüntüden yüksek çözünürlüklü ışık alanı çıkarımı.

Geliştirilen tüm algoritmalar gerçek verisetleri üzerinde test edilmiştir; hem nitel hem de nicel karşılaştırmalar önerilen algoritmaların mevcut literatür çalışmalarından üstünlüğünü göstermek için yapılmıştır.

*Anahtar sözcükler*: uzamsal-değişken görüntü netleştirme, ışık alanı görüntüleme, hesaplamalı görüntüleme, görüntü iyileştirme, yüksek dinamik aralıklı görüntüleme, odak istifleme.

# Acknowledgement

This work would not have been possible without the guidance and support of my supervisor Prof. Bahadır Kürşat Güntürk. Through his constant supervision, I have been able to build up my research work to a stage that I am able to write this thesis. He helped me build a strong foundation of the basic concepts in the area of computer vision. His patience and encouragement has always motivated me to come up with unique ideas to solve challenging problems. I am very thankful to my research committee members Dr. Muhammed Fatih Toy and Dr. Mehmet Kemal Ozdemir for their valuable suggestions on my progress reports throughout the research period.

This thesis is a part of a funded TUBITAK project on light field imaging awarded to Dr. Bahadır Kürşat Güntürk. TUBITAK is a Turkish government organization which funds scientific and research activities throughout the country.

Special thanks to my wife for being a great moral support and for her suggestions about illustrations and graphs in my work. I would like to thanks my parents for their constant moral and financial support. I highly appreciate the suggestions of my lab-partner Umair Mukati, who always spared time for a fruitful research discussion which helped me to overcome some of the hurdles in my research work.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

We are in the midst of an imaging revolution enabled by inexpensive digital cameras and the easy accessibility of the internet. The rapidly growing social media e.g Facebook and Instagram provide a platform to millions of people to share their photos and videos on daily basis. In addition to entertainment, the use of digital imaging finds several applications in a variety of fields, for example, industrial automation, medicine, forensics, surveillance and even vehicle safety. The versatile range of imaging applications, be it the entertainment industry where the visual perception is of prime importance or in artificial intelligence which involves the machine understanding of images, demands high-quality images.

## 1.1 Motivation

In recent times the availability of powerful computational resources has enabled us to fully exploit the potential of image content. For example the use of computational resources, for extracting richer, perhaps more perceptually meaningful information, from either multiple images or conventional cameras with modifications in sensors, optics, and illumination, in the field of computational imaging. Through a variety of computational photography techniques not only many of

the existing limitations of a conventional photography, such as limited dynamic range, poor resolution, limited depth of field, and lens distortions etc can be surpassed but also several new applications of images such as depth extraction and 3D reconstruction, panorama creation, image manipulation (warping, morphing, mosaicing, matting, compositing) and light field imaging can be developed.

## 1.2 Background

A conventional camera typically consists of two parts, the optics and the imager. The optical part of the camera is responsible for converging the light rays reflected by the scene onto the imager. Normally a combination of lenses comprises the optical part of a camera. The role of the imager is to record the intensities of the light rays striking over it surface. The imager is made of light sensitive semi conductor material able to store image in digital format.

Quality of the final image by a typical camera depends on many factors. The ability of the lens to converge the incident light rays onto the sensor determines sharpness of the image. Pixels size and pixel density are two of the major factors that determine an amount of details an image can hold. Sensor's dynamic range is another important factor that effect the image quality as it determines the luminance range of a scene that a camera can capture. In addition to the camera specific factors there are some external factors that also determines the quality of the recorded image, for example, movement of the objects in the scene, a possible camera shake during the acquisition, and atmospheric turbulence etc.

These degradations effect many of the modern days imaging applications. Although some of the camera specific degradations can be eliminated by using high quality optics and sensors designed to capture high dynamic range and high spatial resolutions etc, but these solutions are very expensive and can only solve a specific problem. Through computational imaging techniques a wide range of inexpensive solutions for a variety of image degradations can be developed which

not only improve the existing imaging applications but also open several new possibilities of using high quality images in solving day to day problems. Explained below are some of the possible causes that degrades the images quality, previous solutions to address these problems, their short comings, and some new prospects that can be realized by overcoming these limitations.

#### 1.2.0.1 Space-variant blurring

In many imaging applications, the recorded image is a blurry version of the true image that ideally represents the scene. The common causes of blur are atmospheric distortions, optical aberrations, blur due to averaging on a pixel site, and blur due to motion of the camera and the objects in the scene. The aim of image restoration is to recover the true image from a single or a set of recorded images. The restoration problem is typically ill-posed, requiring regularization techniques to impose certain desired properties on the restored image.

The majority of algorithms addressing the image deblurring problem assume that the degradation process involves a linear shift-invariant blur kernel, which is also known as the point spread function (PSF). When the PSF is unknown, the problem is referred to as blind image deconvolution. Recently, highly successful blind image deconvolution techniques that can handle large motion blur have been developed. The assumption that the entire image is blurred by a single PSF is not valid in general. A non-negligible depth variation in the scene results in space-variant blurring when there is camera shake (i.e., motion blur) or the depth of field is relatively narrow (i.e., defocus blur), as shown in Figure 1.1.

When the scene is static, the blur kernel due to camera shake or defocus is scaled for different depths; therefore, once the kernel is determined, its scaled versions can be used for deblurring different regions in the image. Blur due to rotation of the camera or the scene is another type of space-variant blur whose kernel can be modeled parametrically. In general, the scene can be dynamic where there are objects moving independently in addition to camera movements and optical aberrations. Therefore, it is necessary to handle arbitrary type of

(a) (b) bgh

Figure 1.1: Images blurred by space-invariant and space-variant types of blur. (a) Space-invariant blur (from [1]). (b) Space-variant blur (from [2])

space-variant blur. This is a challenging problem that requires explicit or implicit segmentation of blur regions in the image.

#### 1.2.0.2 Limited dynamic range

Another major limitation of a conventional imaging system is that the sensor′s dynamic range is typically less than the dynamic range of the scene, as shown in Figure 1.2, which limits the expected image quality in digital photography and the performance in machine vision applications. Through capturing and processing multiple images with different exposure values, high dynamic range imaging methods aim to exceed the limited dynamic range. The downside of such an approach is the requirement to capture multiple images, which introduces additional complexities when the scene and the camera are not fixed.

#### 1.2.0.3 Limited depth of field

A common phenomenon that degrades the image quality arises from the limitations of the lens. A typical imaging system, shown in Figure 1.3, has a limited depth of field. For example the rays reflected from a point $p'$ at a distance $d'$ in the scene, converge to approximate a point $p'_r$ on the sensor forming a sharp

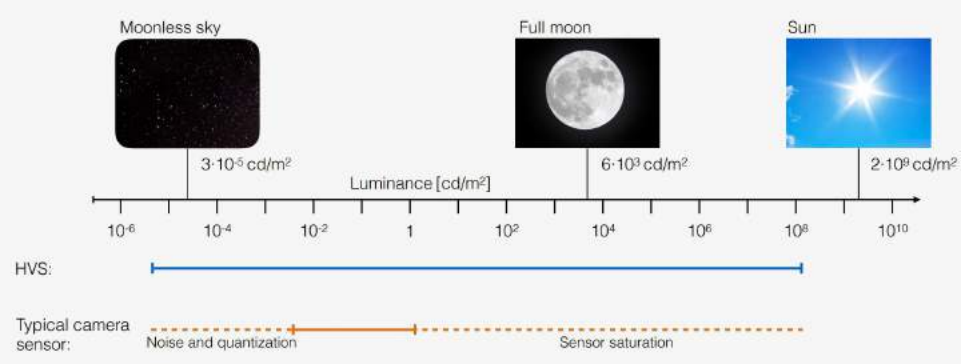



Figure 1.2: luminance range exhibited by the scene compared with the camera's sensor limit to capture the luminance range (from [3]).

image point. However, another point $p$ at a distance $d$ which is relatively farther away from the focus plane reflects the rays with a certain angle that the lens cannot converge these rays to a single point, hence forming a circle of diameter $b_d$ on the image sensor.

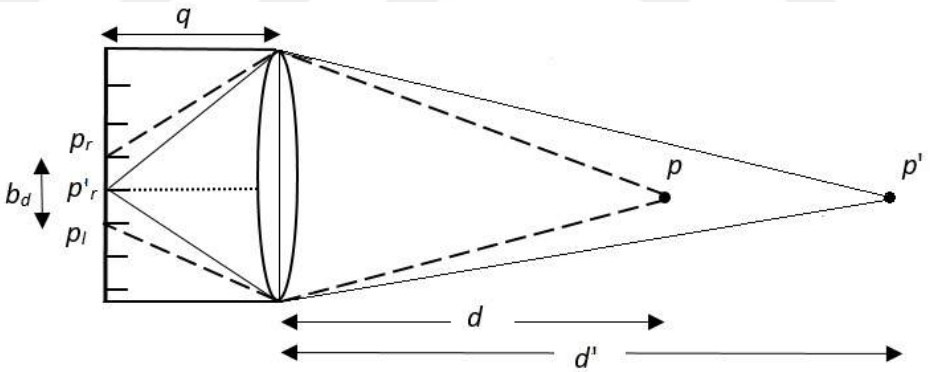
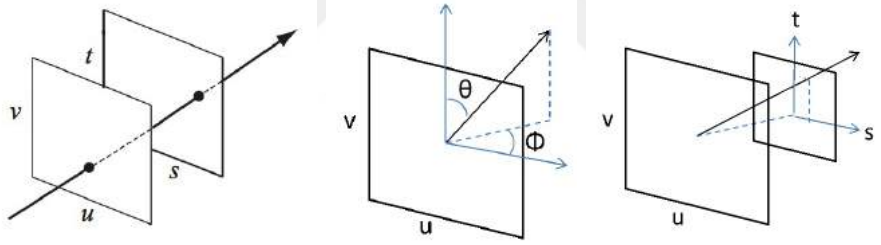



Figure 1.3: Depth of field of a typical imaging system.

Therefore, any object beyond the depth of field of an imaging system begins to defocus. Although shallower depth of field finds many applications in computer vision but from visual perception stand point it is regarded as a degradation. Focus stacking is a common technique in computer vision that aims to extend the depth of field through capturing multiple images of a scene with different focus setting and then combining these images. Again like high dynamic range imaging focus stacking may also suffer from any movement in the scene or the displacement of the camera during the sequential capture.

#### 1.2.0.4 Light field imaging

Apart from overcoming the previously mentioned limitations some new possiblites can be explored through computational imaging techniques, one such example is light field imaging. A light field can be defined as the collection of all light rays in a 3D space [17, 4]. While a light field, in general, can be parameterized in terms of 3D coordinates of ray positions, 2D ray directions, and physical properties of light, such as as wavelength and polarization, the independent parameters can be reduced to a four-dimensional space shown in Figure1.4, assuming there is no energy loss during light propagation and when only the intensity of light is considered; such a four-dimensional representation of light field is used in many practical applications [18, 4, 19].

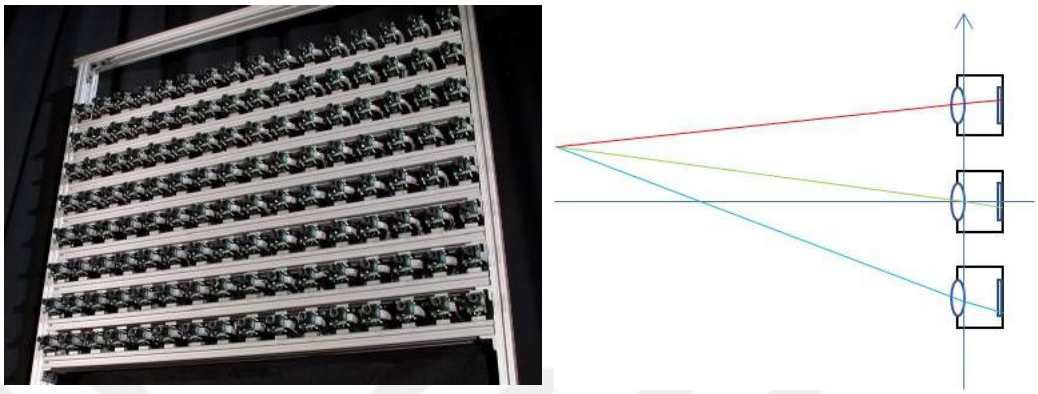



(a) (b) (c)

Figure 1.4: Perimetrized with two parallel plane, in each representation u and v serves as primary arguments. The last two arguments are perimetrized by; (a) Global coordinates of s and t [4].(b) Angular coordinates $\theta$ and $\Phi$ representing the angle of ray after intersecting with uv plane. (c) Local coordinate of s and t, some time also referred as slope of the angle of the ray intersected at uv plane.

#### 1.2.0.5 Light field acquisition

Light field imaging systems can be implemented in variety of ways, including camera arrays [20, 4, 21], micro-lens arrays [6, 22], coded masks [23], objective lens arrays [24], and gantry-based camera systems [25]. Camera arrays configured to capture a light field requires all the camera to be packed tightly as shown in Figure 1.5. Camera array based light field acquisition methods results in high quality light fields as compared to other light field acquisition methods.

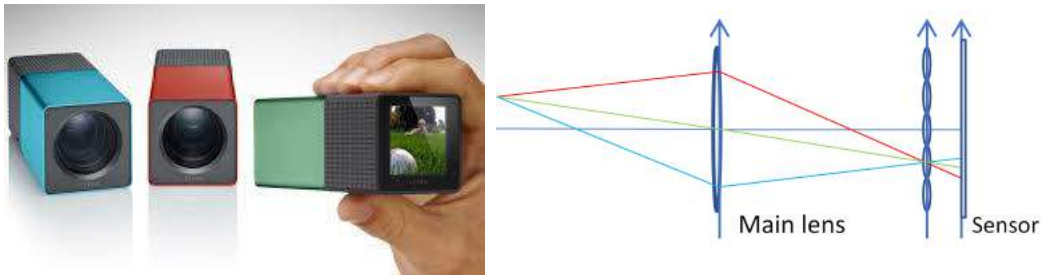


(a) (b)

Figure 1.5: (a) Stanford camera array comprising of 8x16 cameras, configured to capture light field [5]. (b) A point in the scene is projected to three camera sensors of a camera array having independent optical elements.

A camera in an array captures the scene from single perspective on a entire sensor i.e spatial resolution is equal to the camera's sensor resolution and the number of cameras in a camera array determines the angular resolution of the light field. Since all the cameras are triggered simultaneously camera arrays are also suitable for dynamic scenes. The cost to build camera arrays as large as the one shown in Figure 1.5, are very high as compared to any other method of light field acquisition and also they are very bulky and their mobility is restricted.

Among these different implementations, micro-lens array (MLA) based light field cameras offer a cost-effective and compact approach for light field acqusition; and it is widely adopted in academic research as well as in commercial light field cameras [26, 27].




(a) (b)

Figure 1.6: (a) A first generation Lytro camera [6]. (b) Optical diagram of a microlens array based light field camera.

There are many versions of the MLA based light field cameras. Angular and spatial resolution of these cameras depends on the position of MLA in the camera and the size of each micro lens. Lytro which is a commercially available light field camera, shown in Figure 1.6, captures the light field by placing a hexagonal MLA grid at the focal length of the objective lens. Raytrix another type of hand held MLA based design records several micro images from different perspectives by positioning the MLA in front of the sensor which turns the sensor into a micro camera array focused at an intermediate image plane. Irrespective of the MLA position the imaging sensor is shared to capture both spatial and angular information, therefore, MLA-based light field cameras suffer from a fundamental resolution trade-off between spatial and angular resolution, as shown in Figure 1.7.

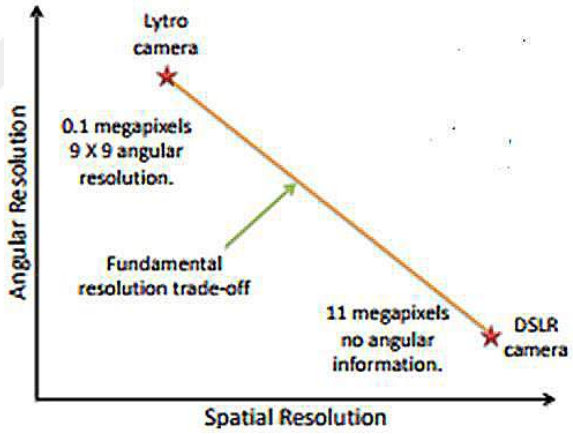



Figure 1.7: Spatial and angular resolution trade-off of an MLA based light field camera. (The illustration is taken from ([7]).

The first-generation Lytro camera has a sensor of around 11 megapixels, producing 11x11 angular resolution and less than 0.15 megapixel spatial resolution [14]. The second-generation Lytro camera has a sensor of 40 megapixels; however, this large resolution capacity translates to only four megapixel spatial resolution (with the manufacturer's decoding software) due to the angular-spatial resolution trade-off.

The second issue associated with MLA-based light field cameras is narrow baseline. The distance between sub-aperture images decoded from a light field

capture is very small, significantly limiting the depth estimation range and accuracy. For instance, the maximum baseline (between the leftmost and rightmost sub-aperture images) of a first-generation Lytro camera is less than a centimeter, which may result in sub-pixel feature disparities. There are methods in the literature specifically designed to estimate disparities and depth maps for MLA-based light field cameras [28, 29, 30].

Coded mask based light field camera design is an alternative to MLA based approach. In mask based approach a coded mask is used to optically modulate the light field prior to projection. There are various design approaches, for example, coded apertures, coded lens arrays, and coded aperture and mask combined. The variation is not only on the placement of the mask but also on the mask design to allow higher permittivity of light. There is also a difference in the design approach based on the number of shots. For example coded mask based camera designs involving multiple-capture [31, 32, 33, 34] and single-capture [8] methods. In Figure 1.8, a coded mask based camera design is shown which requires a single shot to capture the light filed. The mask is placed next to the sensor with a small offset to modulate the light field.

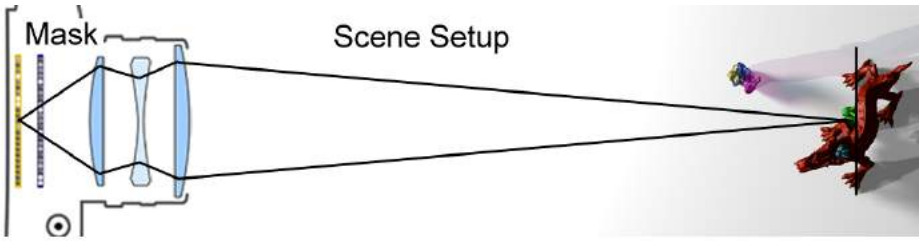


Figure 1.8: Coded mask based light filed camera (The illustration is taken from ([8]).

The limitations of the masked based approach are, either the multiple-capture methods are not suitable to applications with dynamic scenes or single-capture methods may suffer from low light efficiency. Since a single sensor is used to obtain both angular and spatial information, these cameras may also have a low spatial resolution problem, as in the case of MLA based cameras. Gantry based designs do not have a resolution issue however, they are not suitable for dynamic scenes.

#### 1.2.0.6 Light field capabilities

Unlike regular cameras, light field cameras capture the directional light information, which enables new capabilities, including post-capture adjustment of camera parameters (such as focus and aperture size), post-capture change of camera viewpoint, and depth estimation. As a result, light field imaging is getting increasingly used in a variety of areas, including digital photography, microscopy, robotics, and machine vision.

***Post-capture refocusing***

The light field camera captures the angular information of the incident rays separately, which can be utilized to combine different rays computationally to generate a certain perspective image. For example, fixing the angular coordinates $u$ and $v$, i.e picking the pixels from the same $u$ and $v$ location in all the microlenses results in a $u$ x $v$ perspective image. In [6] it has been shown that averaging these perspective images after an appropriate shift produces an image focused at a particular depth in the scene and a significant blur at all the other depths. Figure 1.9, presents an image which is refocused at three different depths computationally.

***Post-capture aperture adjustment***

The aperture size is one of the factors responsible for the quality of an image, as it controls the amount of light that can reach the sensor. Aperture opening is also responsible for the depth of field and hence is very important in many computer vision applications. In Figure 1.10, it is demonstrated that how synthetically changing the aperture size controls the depth of field. The narrow opening allows only a narrow cone angle of the bundle of rays and thus extends the depth of field and vise versa.

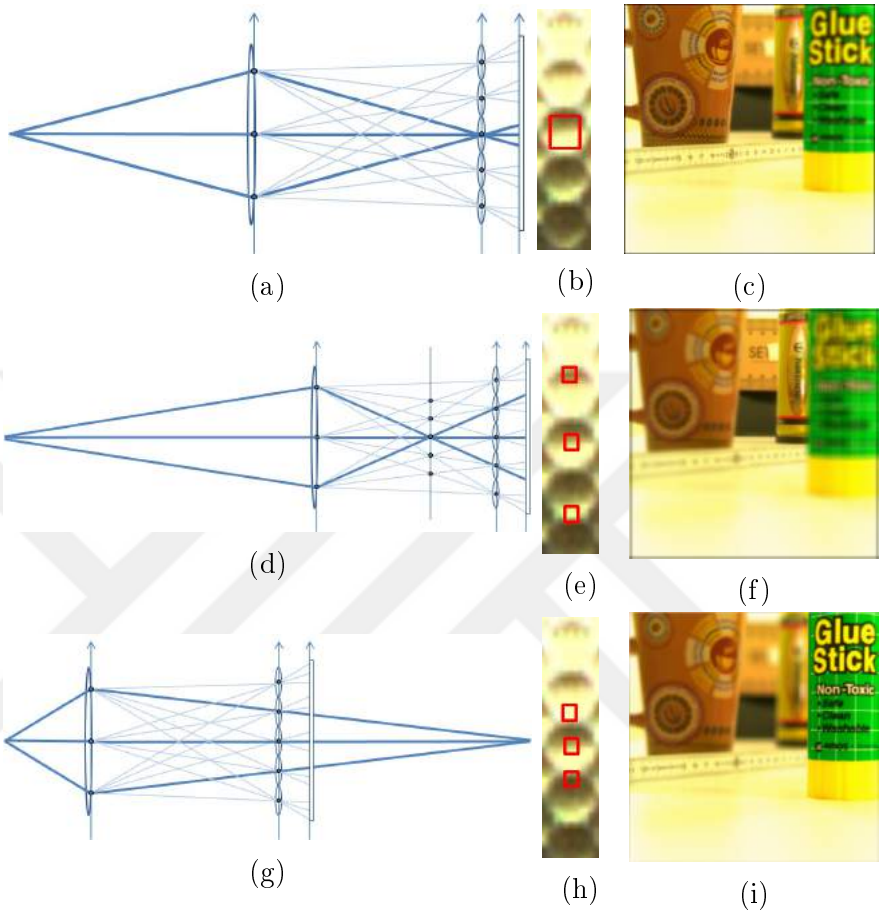



Figure 1.9: (a, d, g) Optical diagram showing converging light rays at virtual image plane; (b, e, h) Pixels picked from raw lenslet marked with red to get refocused image after averaging marked points; (c, f, i) Three refocused images with each image having one depth in focus.

## 1.3 Contribution

This work is concentrated on developing new methods in computational photography to address some of the previously mentioned limitations in generating high-quality images for existing or possibly new applications. The major contributions of this work are as follows:

- A flexible framework for space-variant bling image deblurring is developed.
- A flexible pipeline for joint focus stacked high dynamic range imaging is developed.

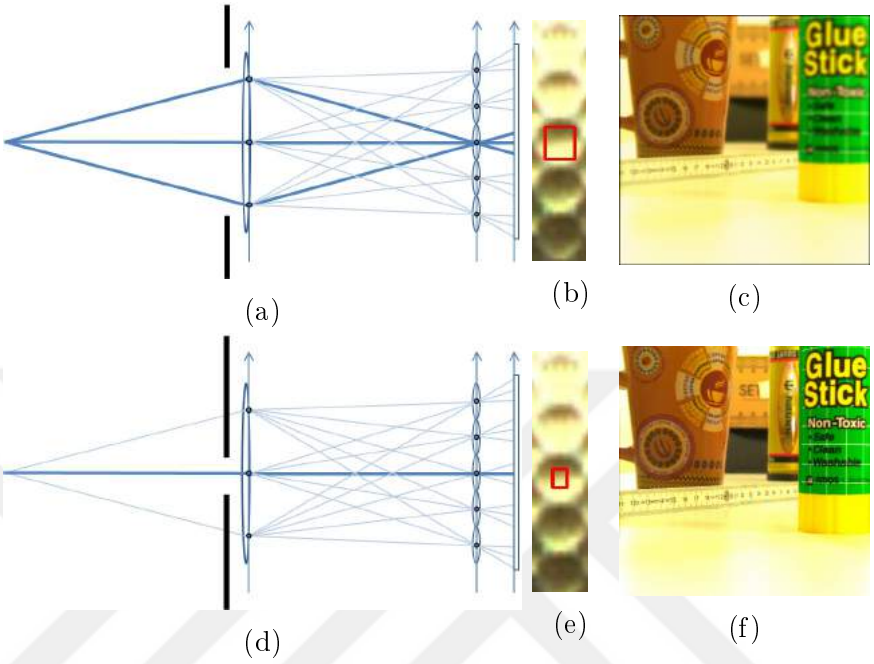


Figure 1.10: (a, d) Optical diagram demonstrating effect of placing a virtual aperture stop; (b, e) Region marked with red square shows the pixel region averaged to get the projected point; (c, f) Reconstructed image.

- An algorithm for Light field spatial resolution enhancement and depth range extension through hybrid stereo imaging is developed
- A prototype light field camera is designed which involves deconvolution to extract high resolution light field extraction from a single capture.

***Flexible framework for space-variant blind image deblurring***

In this thesis a a framework for blind space-variant deblurring without any parametric kernel restrictions is presented. The proposed strategy has the following main steps: coarse blur kernel estimation of small image patches; kernel clustering to determine the prominent blurs in the scene; refining the kernels using the image regions associated with the kernel clusters; and finally space-variant deblurring through deconvolution with the kernel estimates and image fusion. Within the framework, the specific methods (such as blur kernel estimation, clustering, and image fusion) can be altered.

***Dynamic range and depth of field enhancement using camera array***

In this work, a method is presented that combines focus stacking and high dynamic range imaging to generate a final HDR image with extended depth of field using the images captured through a camera array. The first step of the proposed algorithm is to perform focus stacking under exposure diversity. This step involves photometric and geometric registration of the images to produce a set of all in focus images. Radiance map is then estimated using all in focus images with varying exposure followed by tone mapping. The resulting final image has a high dynamic range image and extended depth of field.

***Hybrid Light Field Imaging for Improved Spatial Resolution and Depth Range***

To address both resolution and baseline issues, a hybrid stereo imaging system that consists of a light field camera and a regular camera is proposed. The proposed imaging system has two main advantages over a single light field camera: First, high spatial resolution image captured by the regular camera is fused with low spatial resolution sub-aperture images of the light field camera to enhance the spatial resolution of each sub-aperture image; that is, a high spatial resolution light field is obtained while preserving the angular resolution. Second, the distance between the light field camera and the regular camera produces a larger baseline compared to the maximum baseline of the light field camera; as a result, the hybrid system has a better depth estimation range and accuracy.

***Deconvolution based light field extraction from a single image capture***

The proposed method overcomes the spatio-angular trade-off and enables high-resolution light field capture from a single image. In the experiments, a light fields with 11 x 11 angular resolution and 1024 x 1280 spatial resolution is obtained. The results of the proposed method are compared with light field captured by a first-generation Lytro camera to demonstrate the resolution improvement. In addition to the resolution issue of the MLA based light field cameras, the proposed method also overcomes the poor light efficiency issue of the coded mask based light field cameras. On the downside, the proposed method involves deconvolution, which

may introduce artifacts.

## 1.4 Outline

The thesis addresses a subset of the possible causes of image degradation in a conventional imaging system and explores new possibilities for imaging application by overcoming these degradations. In the next chapter the issue of space-variant blurring in digital images is addressed, the proposed method has been compared with the state of the art methods from the literature. In the third chapter we address two common limitations of the conventional imaging system, low dynamic range and limited depth of field. The proposed algorithm combine focus stacking and HDR imaging to overcome these limitations. Fourth chapter addresses the problem of poor spatial resolution in the context of light field imaging. The developed algorithm employs stereo hybrid imaging that not only overcomes the limited resolution problem but also improves the depth estimation accuracy, which is one of the main applications of light field imaging. In the fifth chapter again the issue of limited spatial resolution in light field imaging is addressed, but this time through a completely different approach. An algorithm is developed to extract a high-resolution light field from a single image through the deconvolution of an input image with a set of pre-estimated PSFs. Finally the thesis is concluded in chapter six.

# Chapter 2

# Space-Variant Blur Kernel Estimation and Image Deblurring through Kernel Clustering

This chapter presents a space-variant blur kernel estimation and image deconvolution framework. For space-variant blur kernel estimation, the input image is divided into small patches, and for each patch, the blur kernel is estimated. The estimated kernels are then grouped to determine different kernel clusters in the image. During clustering, unreliable kernel estimates are eliminated. The blur kernel for each kernel cluster is finally refined using the corresponding image region, which is the union of image patches associated with the kernels in the cluster. For space-variant image deconvolution, the entire image is deconvolved with each blur kernel to produce a set of deblurred images. These images are then fused to produce a blur-free image, where the fusion process selects the optimal regions from the set of deblurred images.

## 2.1 Related work

In this work, a blind image deblurring problem is addressed through only a single image degraded by space-variant blur that does not have any specific parametric form. Since the proposed approach utilizes space-invariant blind deblurring methods in local regions, first a brief of review such techniques, focusing mostly on the ones where the blur kernel is not limited to a parametric form is provided.

A common space-invariant blur type is motion blur due to camera shake during exposure period. The method in [35] requires the user to select a rectangular patch without saturation and an initial guess about the orientation (horizontal/vertical) of the blur kernel. The optimization is done in a Bayesian framework, where the prior of the latent image is a mixture of Gaussians of the gradient image, and the prior of the blur kernel is a mixture of exponential distributions, promoting sparsity of the kernel coefficients. In [36], the blur kernel is also modeled with an exponential distribution, while the latent image model is the product of a global and a local prior. The global prior is defined by modeling the image gradient distribution in the logarithmic space by concatenating two (linear and quadratic) functions. The local prior is defined by modeling the difference between the blurred and unblurred image gradients with a Gaussian distribution. The method requires a rough initial estimate of the blur kernel, which can be in the form of a user-drawn line.

The above-mentioned methods require the user to input an initial estimate of the blur kernel. Alternatively, the initial estimate of the true (unblurred) image can be used. In [37], a "sharp" version of the blurry image is predicted by finding the location and orientation of edges in the image using an edge detector, and then propagating the local maximum and minimum pixel values along the edge profile to form the sharp edges. Once the sharp image is predicted, the kernel estimation is done in a Bayesian framework using Gaussian priors on kernel gradient and noise terms. In [38], the initial unblurred image is estimated by applying bilateral filtering, shock filtering and gradient magnitude thresholding to eliminate small gradients from the estimate. The kernel and latent image

are updated iteratively using least squares estimation technique with L2 norm regularization. During iterations, the gradient magnitude threshold is reduced to include more gradient values in the estimation. The method in [39] also starts with a sharp edge construction with a shock filter, selects significant edges, and makes a coarse kernel estimation using the least squares estimation technique. As a second step, the coarse kernel estimate is refined with selected regions and L1 regularization. Finally, the image is deconvolved with total variation (TV) - L1 method. Alternative to the sparsity promoting total variation and L1-norm regularizations, the use of framelets (for image) and curvelets (for blur kernel) is also investigated [40]. There are also methods where multiple images, with different blur conditions, are used in space-invariant blind image deblurring [41, 42, 43].

Instead of modeling camera shake as a 2D planar motion, it is also possible use 3D camera motion models. In [44], space-variant blur due to 3D rotation (roll, pitch, and yaw) of a camera around its optical center is considered. The parameterized blur model, which is shown to be a linear combination of homographies derived from the 3D rotation, is used to deblur images through approximate marginalization and maximum a posteriori approaches. Alternative to 3D rotation, roll (z-axis rotation) and in-plane (xy) translation model is also used [45]. In [1], instead of applying the homographies to the sharp image, all possible homographies are applied only once to a grid of single-pixel dots to form a blur kernel basis. A space-variant blur due to camera shake can then be generated by linear combination of the kernels in the basis, resulting in an efficient algorithm. There are also hardware-based approaches, where the built-in inertial sensors (gyroscope sensor) of the smart-phones are used for accurate blur kernel estimation [46].

In [47], rotational blur is considered. A transparency map, which is produced by the rotating object or camera, is predicted; the rotational motion parameters are then estimated using transparency map. Another transparency based method is presented in [48], where the blur kernel for the blurred region is obtained using the technique in [35] and the deconvolution is done using the Lucy-Richardson method.

Some space-variant deblurring methods are limited to defocus blur, where the kernel shape is fixed but its scale varies with depth. In [49], image patches with similar content but different defocus blur are determined; and among the patches with similar content, those with less defocus blur are used to deblur the corresponding patches with more defocus blur. The method requires the input image to be segmented according to the defocus level, which is determined by identifying the scale of a Gaussian kernel through optimization. In [50], the segmentation is first done through a local contrast measure. The depth map is then refined through Markov Random Field propagation and graph cut technique. In [51], a blur map (that is, a Gaussian kernel scale map) is generated using a local contrast measure and guided filtering; the input image is then deblurred with different Gaussian kernels using L1-L2 optimization to obtain multiple deconvolved images, which are finally merged into a single image using the blur map. In another method [52], the scale estimation is done using local variance, and the restoration is achieved using the truncated constrained least squares technique. As it turns out, blur kernel scale identification is a critical part of space-variant defocus deblurring. It is also shown that using coded apertures instead of full-open conventional apertures, scale identification performance can be improved [53].

There are methods aiming to first segment images according to blur cues. In [54], a Bayesian classifier, with features including local power spectrum, gradient histogram, maximum saturation and local autocorrelation, is used to segment and image into sharp, linear-motion-blurred and defocus-blurred regions. In [55], the segmentation is done using features obtained from sub-band decomposition indicating likelihood of a small neighborhood being blurred by a candidate blur kernel, which is limited to horizontal and vertical box filters of specific lengths. In [56], it is assumed that there is single blurred object in an unblurred background, and the motion is in a single direction with a constant velocity. The image is first segmented based on image derivatives, the size and direction of the kernel in the blurred region is identified, and finally the blurred region is deconvolved using Richardson-Lucy technique. [9], involves explicitly extracting the depth map, parametric blur kernel estimation and deconvolution for each depth, in an

expectation-maximization framework. In [2], the latent image and motion flow are estimated jointly based on a TV-L1 model. The regularization of the motion flow estimation incorporates edge-map, thus sharp motion boundaries can be maintained. A local blur kernel is defined as a linear kernel whose direction and length is determined by the motion flow.

Patch based kernel estimation and deblurring is also a possible approach. In [57], kernels for local regions are first estimated. Kernels that are adjacent are assumed to have similar blur structure. Earth mover's distance is used to measure the similarity between patch kernels. Finally, the kernels which are correct are taken as such and the incorrect kernels are replaced with their neighboring kernels. In [58], a sharp image is estimated using bilateral and shock filtering; local blur kernels are estimated while enforcing smoothness between neighboring kernels; poor kernel estimates are identified and replaced with neighboring kernels. In [59], initial local blur kernel estimates are revised utilizing both the information of corresponding image regions as well as the correlation with the neighboring kernels.

Recently, convolutional neural networks (CNNs) are used for deblurring of dynamic scenes [60, 11]. In [60], the blurry input image is divided into a set of overlapping patches and then a CNN is used to predict motion blur kernels at patch level. To learn the effective features for predicting motion distributions, a set of candidate motion kernels are generated by discretizing the motion space, i.e., the ranges of length and orientation of the motion vectors. Finally, to fuse the patch level motion kernels into a dense field of motion kernels for the image, a Markov random field model is adopted. The model ensures to choose the motion kernel for each pixel with higher confidence estimated by CNN, and also enforces the smoothness of nearby motion kernels. In [11], a blur kernel free approach is adopted for both blur dataset generation and latent image estimation to avoid kernel estimation related artifacts. To simulate blurring process successive sharp frames are captured with high-speed camera and then integrated over time. For directly recovering the latent image from the blurry input, a CNN model with multi-scale loss is proposed. In the multi-scale loss approach, every intermediate output becomes the sharp image of the corresponding scale and hence enhances

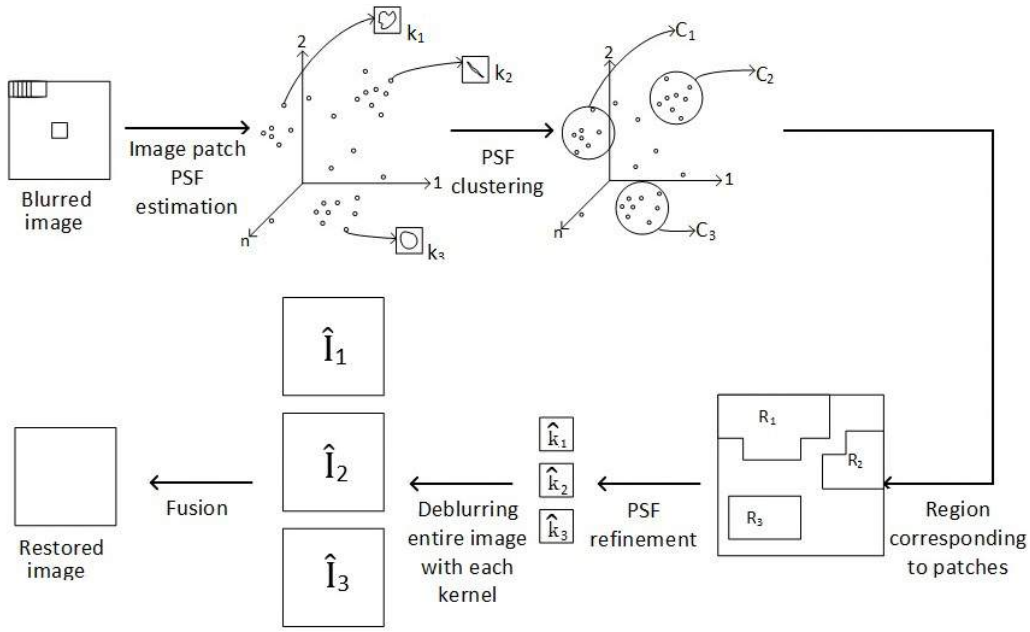



Figure 2.1: An illustration of the proposed space-variant deblurring framework. For each image patch $B_p$, the blur kernel $k_p$ is estimated. The kernels are then clustered to determine the main clusters $C_i$. The image regions $\hat{B}_i$ corresponding to kernel clustered are obtained. For each image region $\hat{B}_i$, refined kernels $\hat{k}_i$ are estimated. The entire input image is deblurred with each kernel to obtain a set of deblurred images $\hat{I}_i$. The deblurred images are then fused to produce a sharp image $I$.

the convergence greatly. These learning based methods can handle multiple types of blur but the effectiveness depends greatly on the diversity of the training data and the success of model fitting.

A single-image blind deconvolution framework where the blur can be space-variant and non-parametric is presented in this chapter. There is no restriction on the type or the number of the blur kernels in the image. The algorithm effectively identifies the blur kernels, segments the image into regions of uniform blur using mean-shift clustering and eventually produces a sharp image through image deblurring and fusion. A comparison of the proposed method with the one given in [61] is presented. A visual and quantitative comparisons with other well-known methods in the literature is also presented.

## 2.2 The blur kernel estimation and deblurring framework

The proposed framework has three main steps: (i) image patch blur kernel estimation, (ii) blur kernel refinement through kernel clustering, and (iii) space-variant deblurring through image fusion. An illustration of these steps is given in Figure 2.1.

### 2.2.1 Image patch blur kernel estimation

The first step is to estimate the local blur kernels in the input image. To estimate the local blur kernel at a pixel location, a small region around that pixel is selected and a space-invariant blur kernel estimation method is applied. In this implementation, the coarse blur kernel estimation method presented in [38] is adopted, which can compute the blur kernels reasonably fast and with acceptable accuracy. The method is applied for each image patch $B_p$ (from the blurry input image $B$) to obtain the latent image $I_p$ and blur kernel $k_p$. The patches do not have to be non-overlapping; and will be detail later, a sliding window approach with a stride of one-fourth of the patch size is used.

### 2.2.2 Blur kernel refinement through kernel clustering

The estimated local blur kernels from the image patches are not all expected to be accurate and reliable. Some image patches may come from transition regions where there are more than one type of kernel; some image patches may come from regions where the texture is not strong enough to produce accurate kernel estimates. It is proposed to perform a clustering process on the estimated blur kernels to determine the major blurs in the image while getting rid of the unreliable kernels. After kernel clustering, the image patches corresponding to a specific kernel cluster are combined to form a larger region for that cluster. Using

larger regions, more accurate kernel estimates are obtained.

With no prior knowledge about the number of blur kernels in the image, the mean shift clustering technique [62] is used. The clustering process starts with a randomly chosen point in the kernel space. The kernel estimates with sum of squared difference less than a fixed bandwidth are added to the cluster. With the added new points, the centroid of the cluster is updated. With the updated centroid, new points are added to the cluster, and the process is repeated until convergence. The clusters are finally checked for possible merging if the distance between two cluster centroids is less than half the bandwidth. Image patches with uniform blur kernel and sufficient texture inside are expected to produce good kernel estimates; and, therefore, they are more likely to form reliable clusters. The patches with non-uniform blur or insufficient texture are expected to have poor kernel estimates, and they do not group with the major clusters.

Once the clusters are obtained, the image is segmented into regions with uniform blur by combining the small patches corresponding to their kernels in the clusters. Kernels $\hat{k}_i$ of the newly combined image regions $\hat{B}_i$ are re-estimated, producing more accurate results. For kernel refinement, the methods presented in [38] and [39] are tested. The method in [39] has higher computational cost than the one in [38] but turned out to produce better results. (For implementation details of the kernel estimators, we refer the readers to [38] and [39].)

In Figure 2.2, a set of input images are provided which are used in the experiments. The figure includes the regions corresponding to kernel clusters and refined kernel estimates $\hat{k}_i$. The patch size and the kernel window size are fixed in all experiments, therefore, different number of kernels are obtained depending on the input image size. (The values of these parameters are given in Section 4.)

### 2.2.3 Space-variant deblurring through image fusion

The clustering process leaves some of the kernels and their corresponding regions un-assigned. These un-assigned patches are extensions of the regions with known

refined kernels, and should be taken care at the end. The Proposed approach is deblur the entire input image $B$ for each refined kernel $\hat{k}_i$ to obtain a set of deblurred images $\hat{I}_i$, and to apply an image fusion process, picking the best pixel value at a pixel position among the alternatives coming from the deblurred images $\hat{I}_i$, to produce a sharp image $I$.

For image deconvolution, TV-L1 based method [63] is used, which minimizes the following cost function $\|B - \hat{k}_i \otimes \hat{I}_i\|_1 + \lambda \|\nabla \hat{I}_i\|_2$, where $\lambda$ is a regularization constant.

For the fusion process, at each pixel position, one pixel value from one of the deblurred images is selected. The selection is based on the observation that in a restored image $\hat{I}_i$, regions for which the applied kernel is correct, a smooth recovery is achieve; while severe ringing artifacts appear in the rest of the image because those regions cannot be explained by that kernel. These ringings are intense and have gradients with large magnitude compared to the gradients of a natural image. This cue has been used before to determine the scale of a blur kernel for depth from defocus [53]. This cue is used in the fusion process in addition to the sparse distribution of gradients in a natural image. In the implementation, a pixel is selected from the deblurred image which produces the minimum energy in a local window around that pixel. The energy function to be minimized is defined as:

$$\sum_{(x,y)\in W} |B(x,y) - \hat{k}_i \otimes \hat{I}_i(x,y)|^2 + \alpha |\nabla \hat{I}_i(x,y)|, \tag{2.1}$$

where $(x, y)$ is a pixel position, $W$ is a local window around that pixel and $\alpha$ is a regularization constant.

## 2.3 Experimental results

the proposed framework is tested on a variety of images degraded with space-variant blur of different types. The comparisons with the following methods, which can handle space-variant blur are done:

- Joshi et al. (2010) [10]
- Hirsch et al. (2011) [1]
- Shen et al. (2012) [51]
- Hu et al. (2014) [9]
- Qian et al. (2014) [61]
- Sun et al. (2015) [60]
- Nah et al. (2017) [11]
- Shen et al. (2018) [57]

From Figure 2.3 to 2.10, a visual comparisons is provided. In Table I,a quantitative comparison using a dataset from [12] is given.

The proposed method has several parameters whose values should be set. One of them is the patch size. The patch size be carefully chosen; and it should be greater than the size of the blur kernel. On one hand, when the patch size is too large, the region may contain more than one type of blur, which would reduce the accuracy of the kernel estimate. On the other hand, a small patch size may result in inaccurate estimates due to insufficient texture within a patch. In all the experiments, the patch size is set to 312 x 208, which is one-sixth of the size of the images that is initially worked with and used to decide on the parameter values empirically.

Another parameter is the stride amount in the sliding window for obtaining the local image patches. If a sliding window with one pixel stride is used, the computational cost would be maximum since the kernel for each patch is going to be estimated. If non-overlapping patches are used, regions of uniform blur could be missed. In all experiments, the stride in horizontal and vertical directions is set to one-fourth of the patch size.

The kernel clustering step employs the mean-shift clustering algorithm, for which the distance measure is sum of squared difference between two kernels,

and the bandwidth is set to 4.5, which is determined empirically. For kernel estimation, the parameters are set to the recommended ones in the paper [39]; and for image deconvolution, again the recommended parameters in [63] are used. For image fusion, the local window size is set to the size of the blur kernel size, and the regularization constant $\alpha$ is set to 10. Once decided, the parameter values are kept fixed in all the experiments.

In Figure 2.3, the input image involves defocus blur and motion blur in at least two different directions as can be seen in the patch blur kernels in Figure 2.2. A method designed for space-invariant blur, such as the one in [39], cannot work well with this type of an image. A method designed for defocus blur, such as [51], also fails to get rid of motion blur. The proposed method in this case works the best. It should be noted that the kernel refinement of the proposed method is based on [39], which is indeed producing very good results when the blur is space-invariant, but fails otherwise as seen in this example.

The input in Figure 2.4 has blur due to camera shake; the scene is also non-planar. The input image is taken from [1]. It is seen that the proposed method works better than [1] as well as [61].

The input in Figure 2.5 has objects at different depths, and there is motion blur, while it is not as much as the one in Figure 2.4. The input image is taken from [9]. The proposed method produces sharpest result compared to [9] and [61].

The input image in Figure 2.6 is taken from [57]. The proposed method is compared with [57], [10], and [61]. The proposed method has a better recovery overall as seen in different close-up regions.

Recently developed convolutional neural network (CNN) based methods can produce exceptional results with space-variant blur. The input images in Figures 2.7, 2.8, and 2.9 are taken from [11], which is such a CNN based space-variant deblurring method. In these figures, results of [2] and [60], which are also designed for removing space-variant blur are included. Evaluating these results, it can be

said that the proposed method produces better or comparable results with all these methods.

Finally, a quantitatively comparison of the CNN based methods [11] and [60] with the proposed method is presented. The dataset is taken from [12], which provides ground truth images. Table I, provides peak signal to noise ratio (PSNR) and structural similarity (SSIM) values for a number of images with various amounts of blur. More information on these datasets can be found [12]. In Figure 2.10, a sample visual comparison from this dataset is presented.

It is seen that while some methods are designed for specific types of blur situations, the proposed method works with a variety of blur situations, producing comparable or better results compared to the other methods, even under conditions where the competing method is specifically designed for such a scenario.

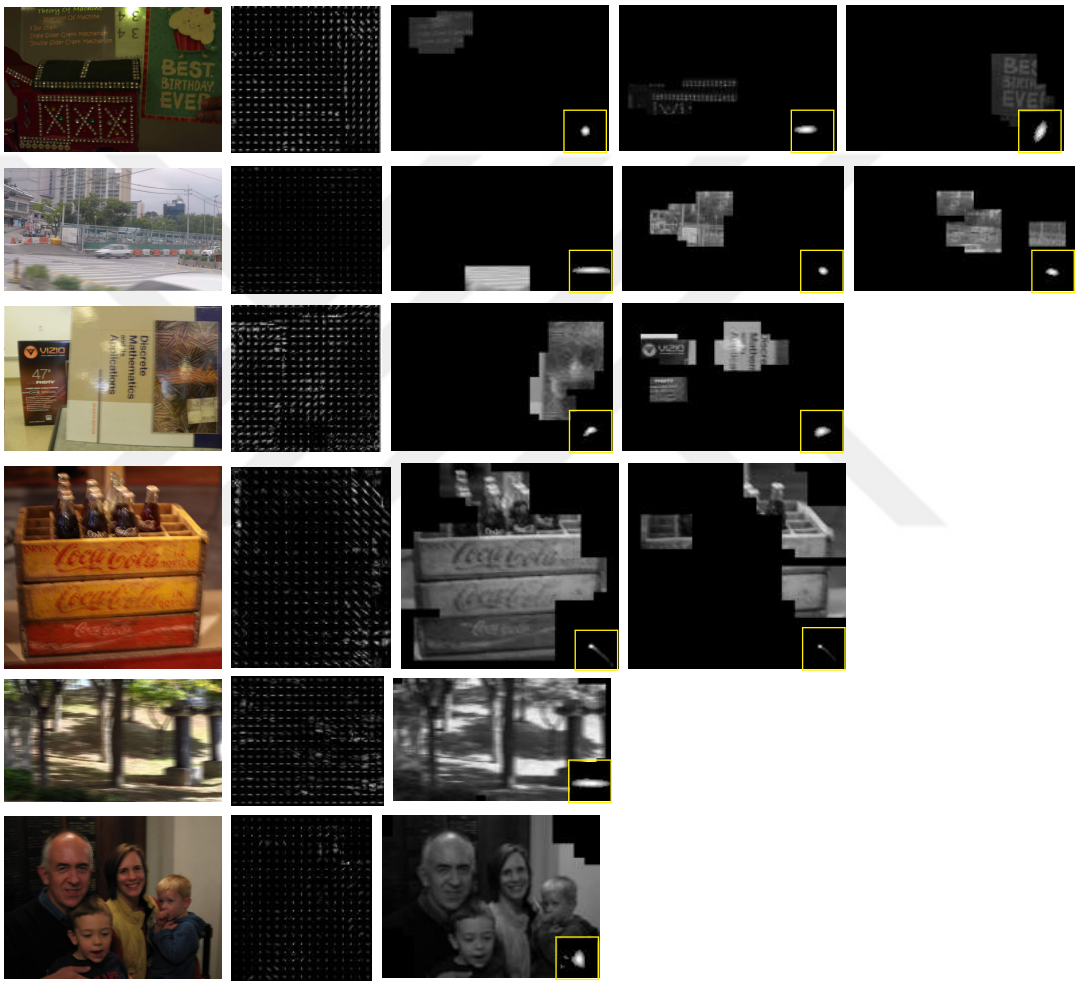


Figure 2.2: Blur kernel estimation and clustering. (From left to right) Original image, kernels estimated for image patches, and image regions after kernel clustering along with the corresponding refined kernels. The patch size and the kernel window size are fixed in all experiments, therefore, depending on the image size, different number of kernels are obtained. The input images and the kernel clusters are scaled to fit into the figure.


Figure 2.3: Comparison of space-variant image deblurring methods.

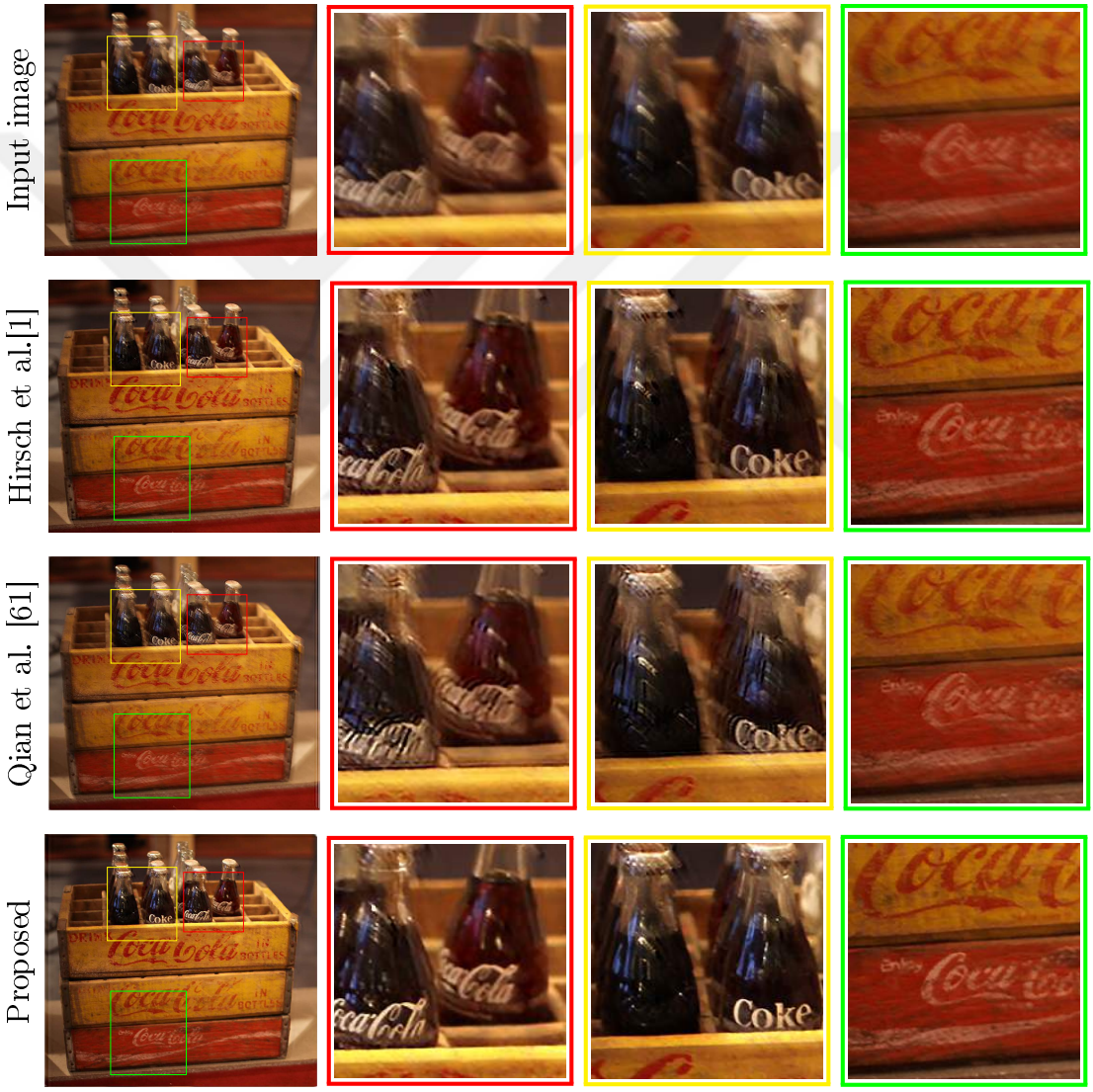



Figure 2.4: Comparison of space-variant image deblurring methods. The input image is taken from [1].


Figure 2.5: Comparison of space-variant image deblurring methods. The input image is taken from [9].

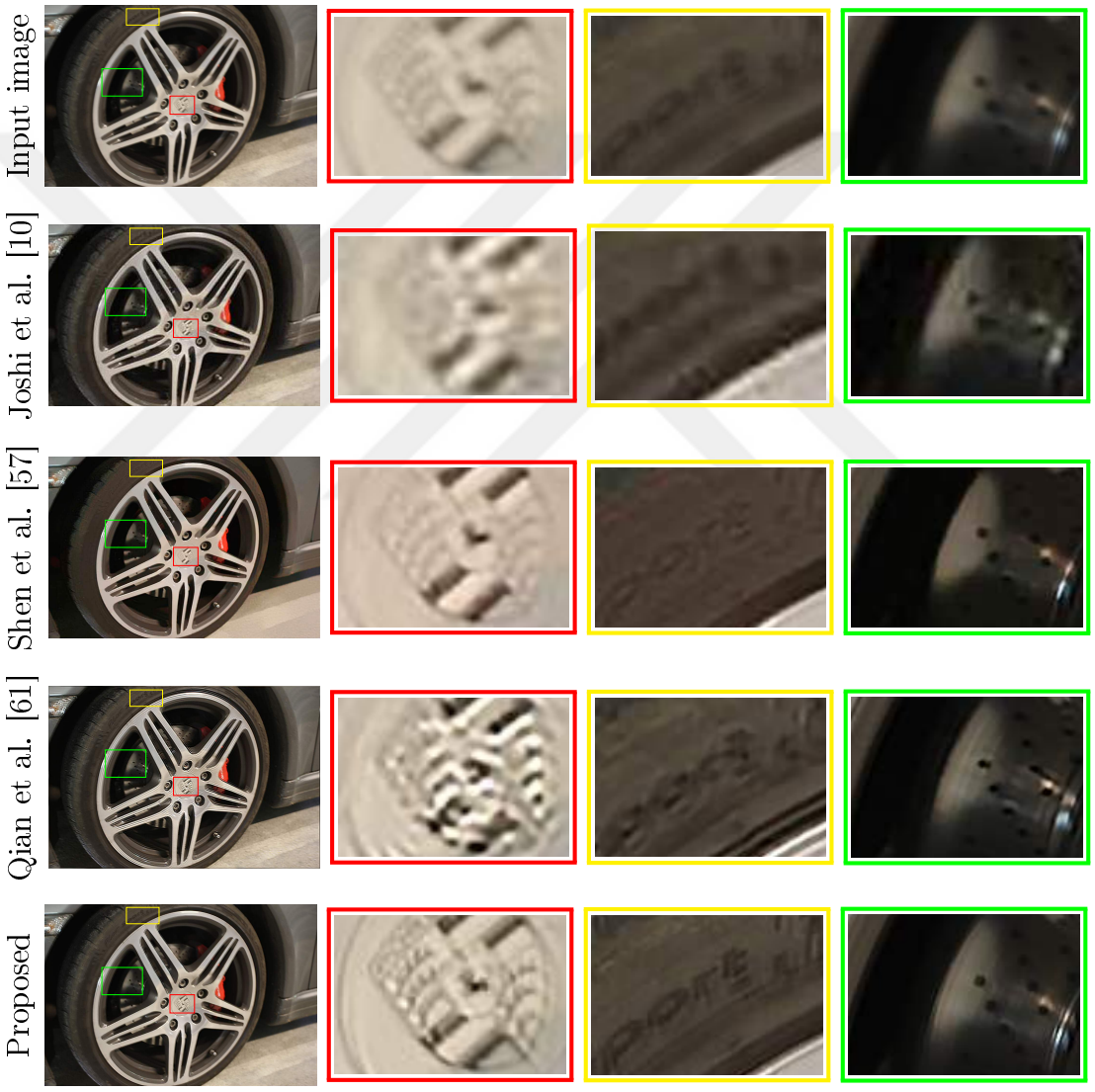



Figure 2.6: Comparison of space-variant image deblurring methods. The input image is taken from [10].

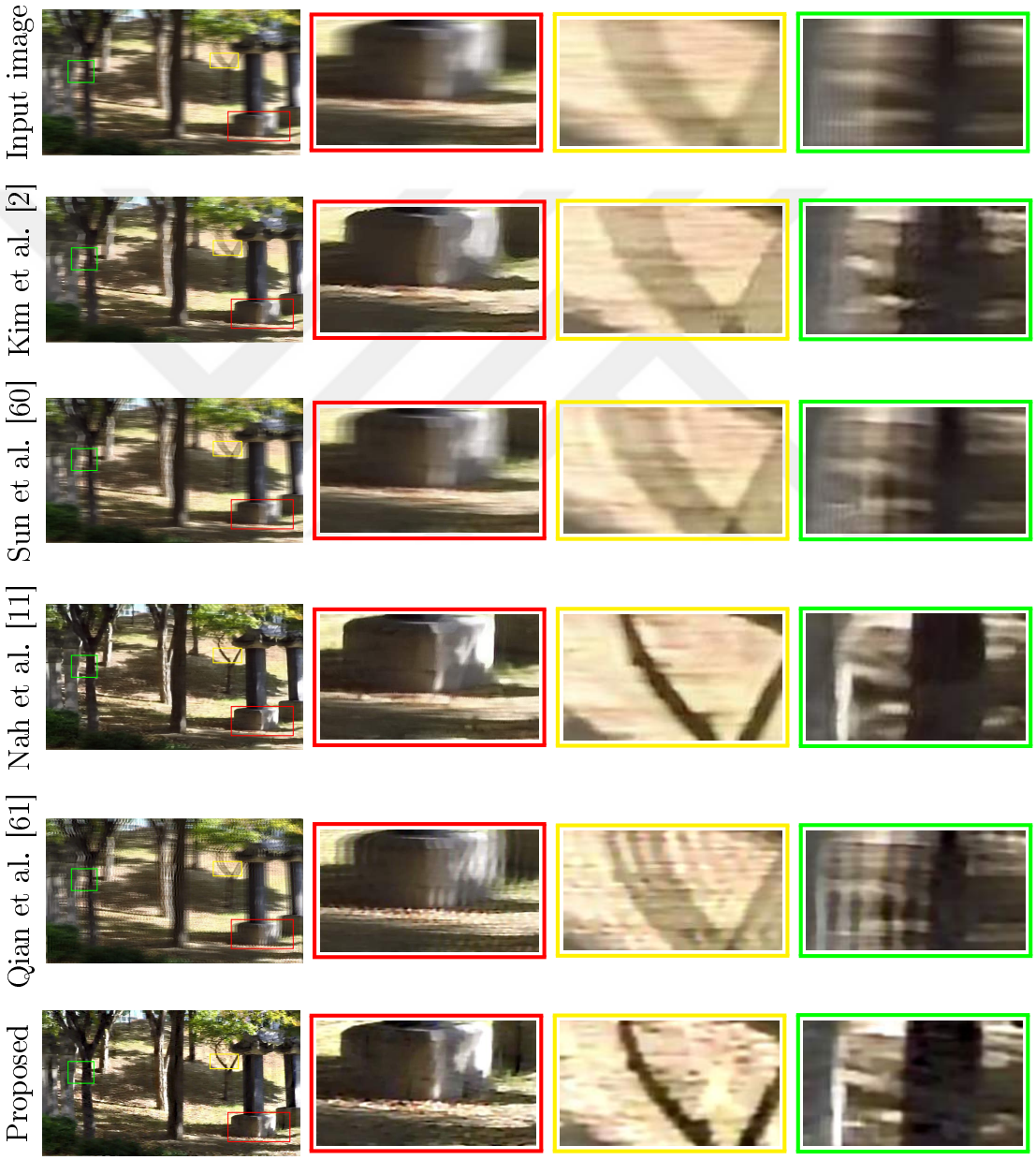



Figure 2.7: Comparison of space-variant image deblurring methods. The input image is taken from [11].

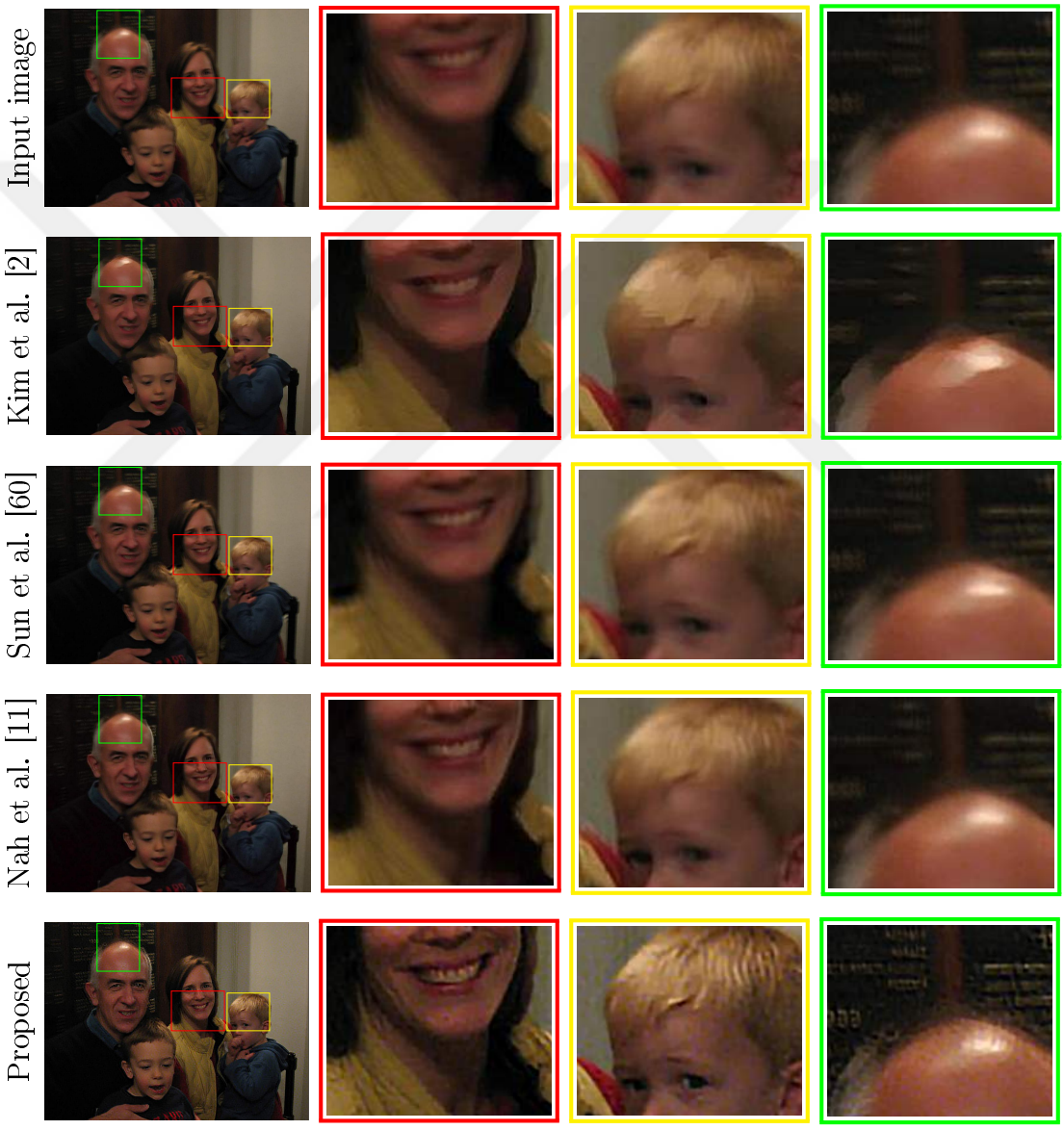



Figure 2.8: Comparison of space-variant image deblurring methods. The input image is taken from [11].


Figure 2.9: Comparison of space-variant image deblurring methods. The input image is taken from [11].

Table 2.1: Quantitative comparison of the proposed method with some other space-variant image deblurring algorithms. The dataset is taken from [12].

| Image Name | Sun et al. [60] | | Nah et al. [11] | | Proposed | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| *Manmade 01* | 17.51 | 0.75 | 16.54 | 0.71 | 18.41 | 0.79 |
| *Manmade 02* | 14.16 | 0.61 | 14.15 | 0.59 | 15.62 | 0.67 |
| *Manmade 03* | 16.82 | 0.72 | 18.12 | 0.77 | 21.25 | 0.87 |
| *Manmade 04* | 16.72 | 0.71 | 19.33 | 0.80 | 20.16 | 0.84 |
| *Natural 01* | 20.42 | 0.87 | 20.62 | 0.87 | 20.26 | 0.86 |
| *Natural 02* | 17.77 | 0.79 | 18.35 | 0.81 | 19.61 | 0.84 |
| *Natural 03* | 18.82 | 0.83 | 19.41 | 0.85 | 23.09 | 0.93 |
| *Natural 04* | 19.20 | 0.83 | 19.98 | 0.86 | 20.14 | 0.86 |
| *People 01* | 31.48 | 0.96 | 32.29 | 0.95 | 31.09 | 0.96 |
| *People 02* | 27.36 | 0.91 | 28.01 | 0.91 | 27.44 | 0.92 |
| *People 03* | 31.02 | 0.95 | 31.11 | 0.94 | 34.10 | 0.98 |
| *People 04* | 31.10 | 0.95 | 31.99 | 0.95 | 32.37 | 0.96 |
| **Average** | 21.86 | 0.82 | 22.49 | 0.83 | 23.62 | 0.87 |

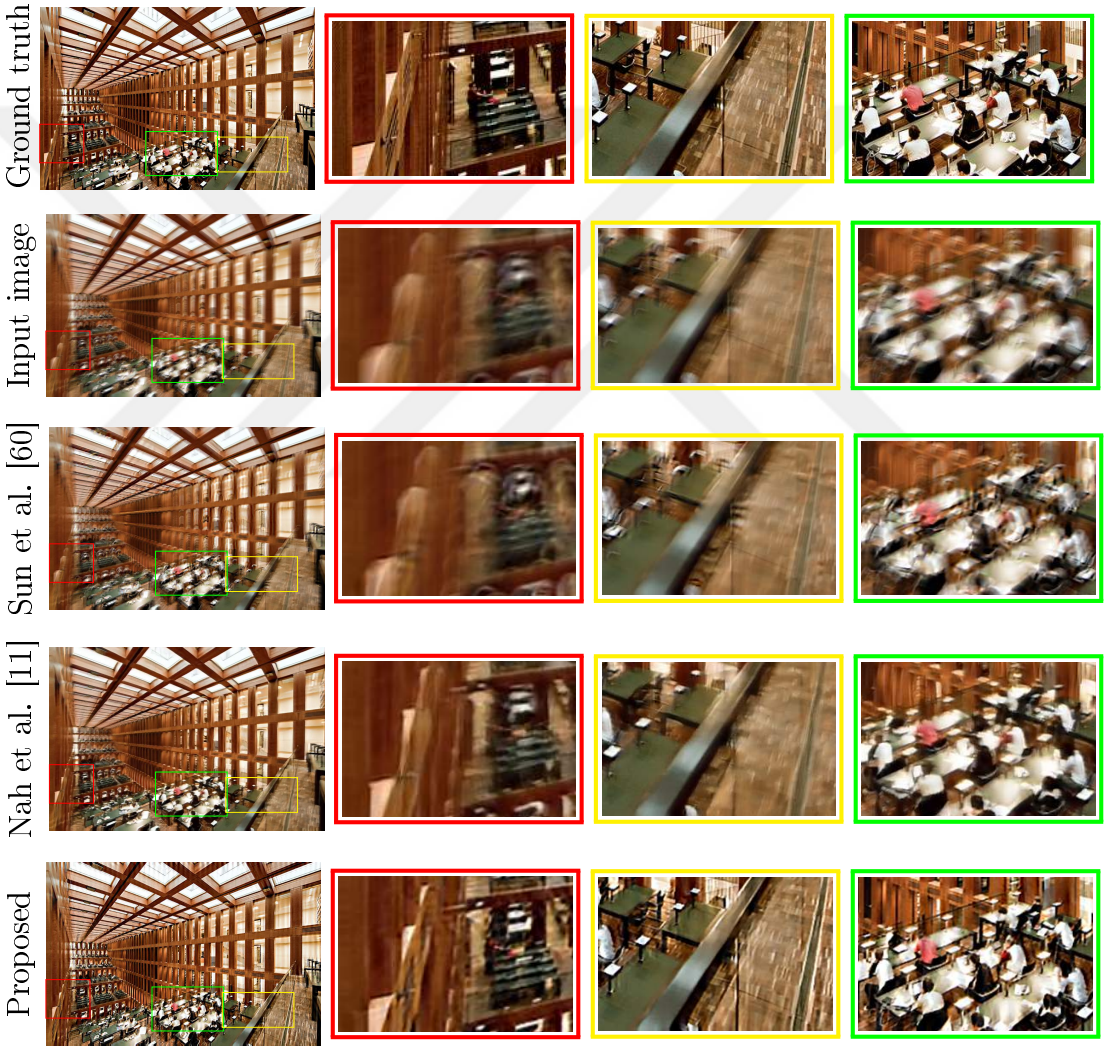



Figure 2.10: Comparison of space-variant image deblurring methods. The input image is taken from [12].

## 2.4 Discussion

The proposed framework involves multiple steps, and the Table 2.2 provides a breakdown of the computational time for each step. Since the number of clusters and the size of clustered regions depend on the input image, there is no fixed time for the entire process. In the table, an average computation times (10 runs) from the dataset of images with size 1248 x 1872 is presented. The patch size is set to 312 x 208, and the stride amount is set to 1/4 of the patch size in each run. The experiments are done on a machine with Intel Xeon(R) CPU @ 3.50 GHz and a memory of 4.7 GB. It should be noted that some of the steps can easily be parallelized: image patches can be processed in parallel to get the initial blur kernels, kernel refinement for each cluster can be done in parallel, and image deblurring for each kernel can be done in parallel.

Table 2.2: Computational time for each step involved in the proposed space-variant image deblurring framework

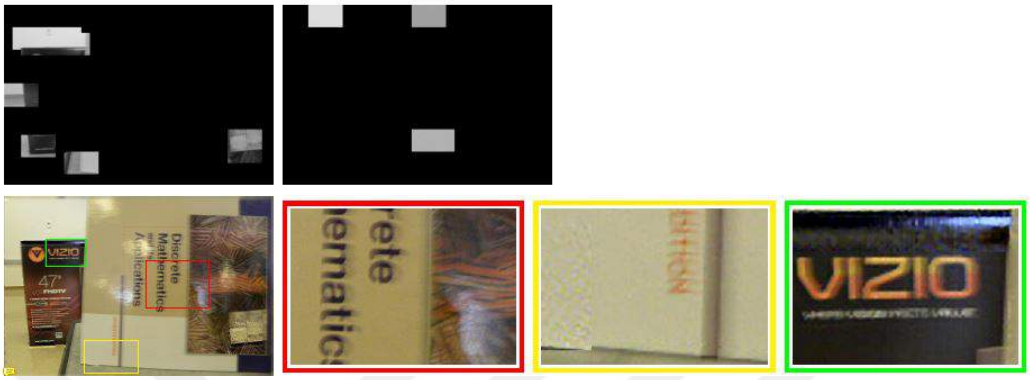
| Computational Time | |
|---|---|
| *Algorithm step* | *Average time* |
| Image patch blur kernel estimation | 4 secs per patch |
| Kernel clustering | 2 secs |
| Kernel refinement | 6 secs per image |
| Image deblurring | 70 secs per image |
| Image fusion | 26 secs |

As already mentioned, the proposed framework is flexible to accommodate changes in the techniques used at different steps. The clustering, kernel estimation, deblurring, or fusion techniques can be replaced with others. The parameters of the techniques can be fine-tuned to obtain the best overall performance. In all experiments that are reported, a a fixed set of parameters are. These parameter values have been determined through experimental evaluation. Here, a further discuss of the effect of these parameters on the performance is provided. One of the critical parameters is the patch size. When it is too small, there may not be enough image features to obtain a reliable blur kernel; when it is too large, it may not be possible to have image patches covering only one type of blur. In Figure 2.11, three patch sizes are compared. The patch size 312x208 results

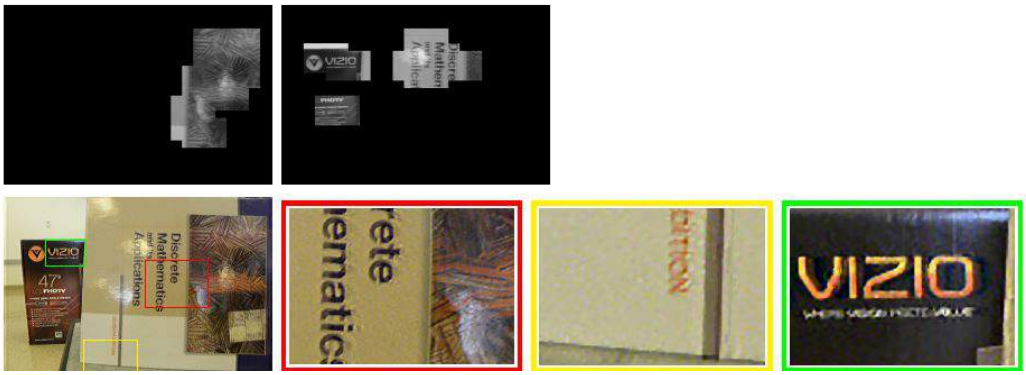
in sufficiently large cluster regions with uniform blur within each cluster. The smaller patch size results in many small cluster regions, while the larger patch size results in a large cluster region, which incorporates more than one type of blurring; and in both of these cases, the deblurring performance degrades. The second critical parameter is the stride amount, indicating shift from one image patch to another. In the extreme case, the stride amount would be set to one pixel; this would increase the number of image patches (therefore, the number of estimated blur kernels), and it is expected to have reliable kernel clustering. On the downside, the computational cost would increase since the blur kernel for each patch has to be estimated. As the stride amount is increased, it is expected to get performance degradation as the number of kernels is decreased and the chance of missing uniform kernel regions is increased.

In Figure 2.12, the stride amounts of 1/4 and 1/2 of the patch size are compared. It is seen that the larger stride amount (1/2 of the patch size) results in visible artifacts in the final deblurred image.

Another important parameter is the mean-shift clustering bandwidth. When the bandwidth is too large, different types of kernels may be joined together; when the bandwidth is too small, kernels (even from the same type of blur) cannot be grouped and some blur types may be skipped. In Figure 2.13, results with different bandwidth values are provided; it can be seen that the bandwidth value of 4.5 produces a good result compared to the others. Finally, whether the initial point in kernel clustering has any significant effect on the performance is investigated. The seed point is selected randomly and repeated the process ten times. As expected, different seed points may result in slightly different clusters; however, this does not result in any significant variation in the final deblurred image. This is exemplified in Figure 2.14.



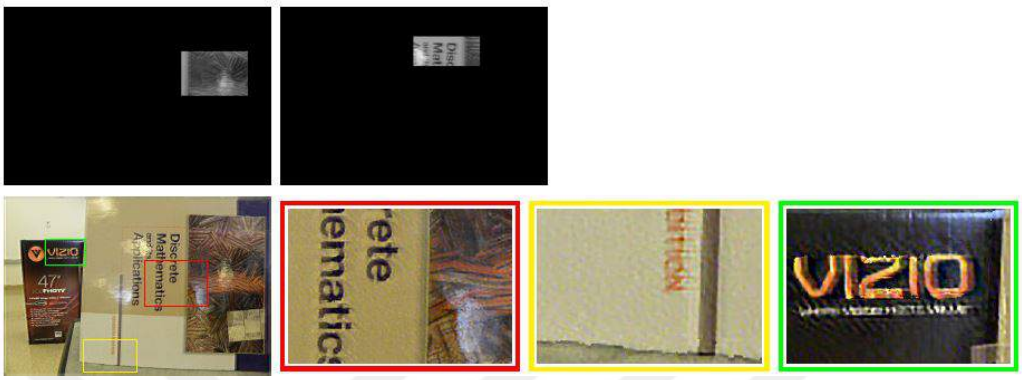

(a) The patch size is 240 x 156. (Top) Clustered regions. (Bottom) Final deblurred image and zoomed-in regions.




(b) The patch size is 312 x 208. (Top) Clustered regions. (Bottom) Final deblurred image and zoomed-in regions.

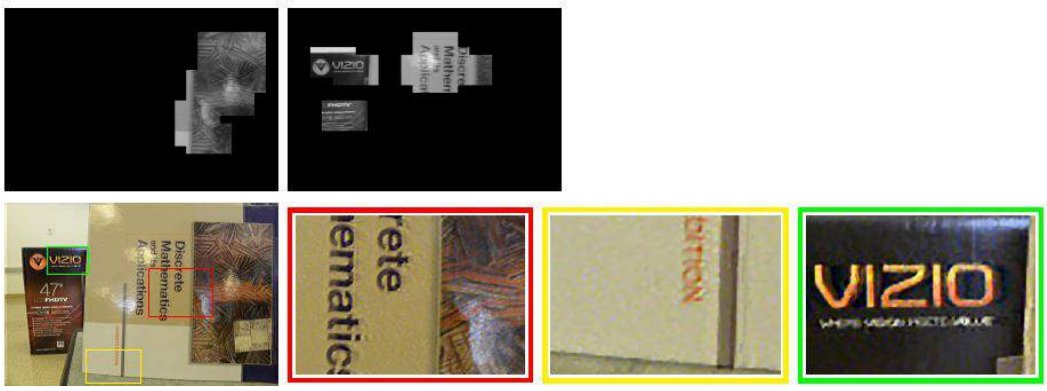

(c) (The patch size is 468 x 312. (Top) Clustered regions. (Bottom) Final deblurred image and zoomed-in regions.

Figure 2.11: Effect of patch size on clustering and final deblurring. Clustered regions and deblurred images are shown for different patch sizes. The stride amount is 1/4 of the patch size in each case.




(a) The stride amount is 1/2 of patch size. (Top) Clustered regions. (Bottom) Final deblurred image and zoomed-in regions.

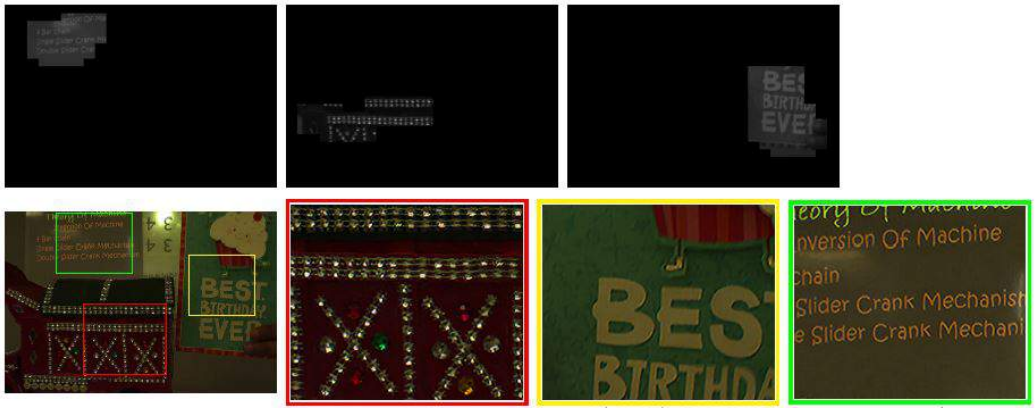

(b) The stride amount is 1/3 of patch size. (Top) Clustered regions. (Bottom) Final deblurred image and zoomed-in regions.




(c) The stride amount is 1/4 of patch size. (Top) Clustered regions.(Bottom) Final deblurred image and zoomed-in regions.

Figure 2.12: Effect of stride amount on clustering and final deblurring. The patch size in each case is 312 x 208.

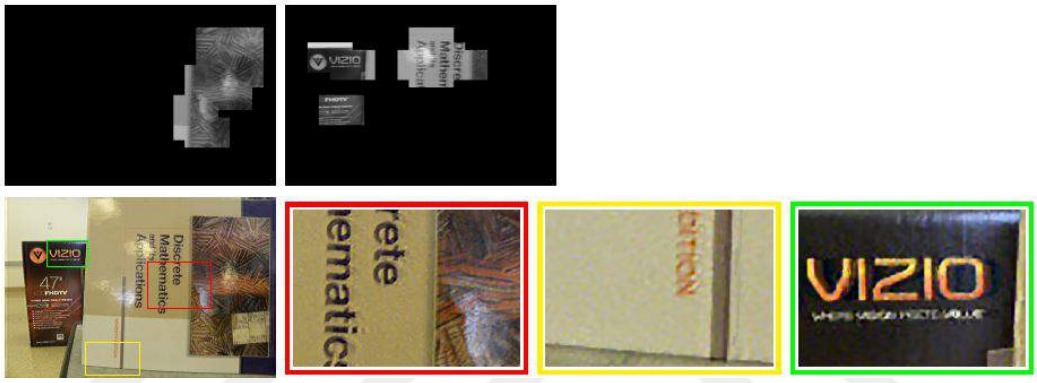
(a) The mean shift clustering bandwidth is 4.0. (Top) Clustered regions. (Bottom) Final
deblurred image and zoomed-in regions.



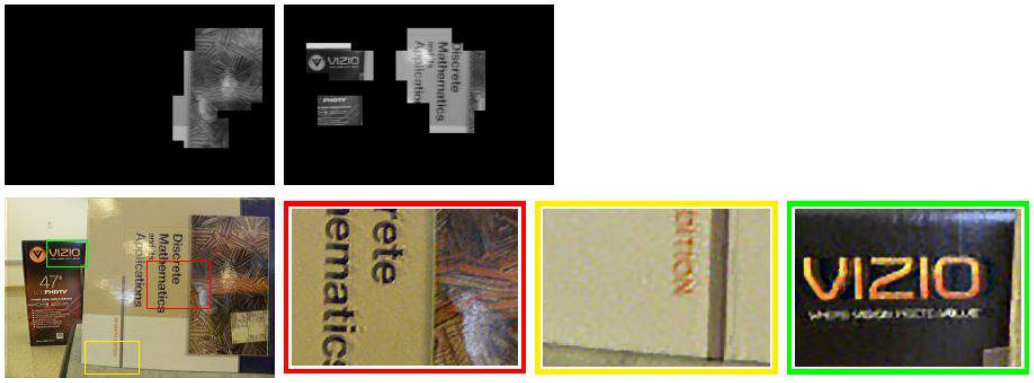
(b) The mean shift clustering bandwidth is 4.5. (Top) Clustered regions. (Bottom) Final
deblurred image and zoomed-in regions.

(c) The mean-shift clustering bandwidth is 5.0. (Top) Clustered regions. (Bottom) Final
deblurred image and zoomed-in regions.

Figure 2.13: Effect of mean-shift clustering bandwidth on clustering and final deblurring.




(a) (Top) Converged regions. (Bottom) Final deblurred image and zoomed-in regions




(b) (Top) Converged regions. (Bottom) Final deblurred image and zoomed-in regions.

Figure 2.14: Random selection of clustering seed points and its effect on the clustering convergence and final deblurring. The result in (a) is obtained in 7 out of 10 runs; the result in (b) is obtained in 3 out of 10 runs.

# Chapter 3

# Dynamic Range and Depth of Field Extension using Camera Array

Combining multiple digital cameras in different configurations and forming large camera arrays enable several new imaging capabilities. Cameras separated by a small spacing and arranged such that their field of view is overlapping, allows simulating a wider synthetic aperture. Such a wider aperture can be useful to see through in partially occluded environments. Camera array with narrower spacing between the cameras and staggered triggering times create a virtual high-speed camera. Varying the exposure of cameras in a camera arrays enables an interesting possibility i.e high dynamic range imaging.

The configuration of the camera array in this work helps to overcome the limitations of traditional cameras in terms of their dynamic range and depth of field. In this chapter we present a method that combines focus stacking and high dynamic range imaging to generate a final HDR image with extended depth of field, using the images captured through the camera array. The first step of the proposed algorithm is to perform focus stacking under exposure diversity. This step involves photometric and geometric registration of the images to produce a set of all in focus images. Radiance map is then estimated using all in focus images with varying exposure followed by tone mapping. The resulting final

image has a high dynamic range image and an extended depth of field

## 3.1 Related work

Camera arrays have been used in the past for different imaging applications such as, high dynamic range imaging, high-resolution imaging [5], high-speed video [64], reconstructing occluded surfaces [65], [66], and blur removal in [67]. In [5] the cameras are arranged as 12 x 8 array with around 50% overlapping field of view between adjacent cameras. Autostitch software is used to perform feature detection and correspondence, bundle adjustment, and generate HDR mosaiced image. To capture a high-speed video a total of 52 cameras are packed closely together to approximate a single center of projection in [64]. They achieved a higher frame rate by staggering the exposure time of each camera and to align the frames from each camera to a reference view the scene is assumed to be planar and hence projective transformation is used.

In [65] a total of 88 cameras were used to generate a large synthetic aperture. By synthetically focusing the large virtual lens on objects behind occlusion, the defocus blur of occlusion projections becomes obvious and hence the occlusion does not appear in the image. In another attempt to see through occlusion, the visual parallax of the objects at different depths and the sub-pixel information of the images captured by the camera array is used in [66] to determine the shape of occlusion and regenerate the background. To address the problem of spatially varying blur in images and videos a hybrid camera array is proposed in [67]. The camera array captures high-resolution low frame rate images and low-resolution high frame rate images. These images are fused to generate a deblurred high-resolution high frame rate video. In this work, the camera array is diversified such that it has variations in both the exposure time as well as in the depth of field. The change of exposure enables room for high dynamic range imaging and the variation of camera's focus in the array allows for extending the overall depth of field.

High dynamic range imaging has been studied extensively and several methods have been proposed to extend the dynamic range. These methods can be categorized into two types, sensor design methods, and multi-exposure capture. In [68] an optical mask with spatially varying transmittance is placed adjacent to the sensor to change the exposure of adjacent pixels. A modified camera sensor is used in [69] and [70] where the sensor design includes detector cells with different size sensing elements to record different measurements in each cell which are then combined on-chip to generate HDR image. In [71] a camera design has been proposed that measures alternating 'cliques' of the sensor to locally determine the best exposure and use Poisson solver to reconstruct the HDR image.

HDR imaging algorithms that use multi-exposure images involve four basic steps. 1) Multiple exposure images acquisitions. 2) A camera response curve estimation. 3) Applying the inverse camera response curve on the images. 4) Merging the linearized images. There are several different ways of capturing multi-exposure images [72], [5], [73], [74] as well as camera response function (CRF) estimation [72], [75], and [76]. In [72] a set of images of a static scene are captured sequentially with different exposure times by a single camera from a fixed location. A large size camera array is used in [5] to capture a dynamic scene from multiple cameras with different exposure settings. In [73] and [74] light field cameras are used to capture multi-exposure images. In [73] an optical mask with higher transmittance in the middle and lower transmittance at the border is introduced in front of the main lens to increase the effect of vignetting resulting in multiple exposure sub-aperture images. Plenptic 2.0 camera is used in [74] to capture multiple-exposure microcamera images. The change in the size of the aperture of a microlens based on its f-number results in the change of exposure of each microlens image. There is also a wide range of camera response function estimation models in the HDR imaging literature, while [72] approximates the CRF by a simple gamma function [75] proposes a non-parametric model and defines the curve by a set of exposure values for a better approximation. In [76] the camera response curve is expressed by a low order polynomial to provide a flexible radiometric model.

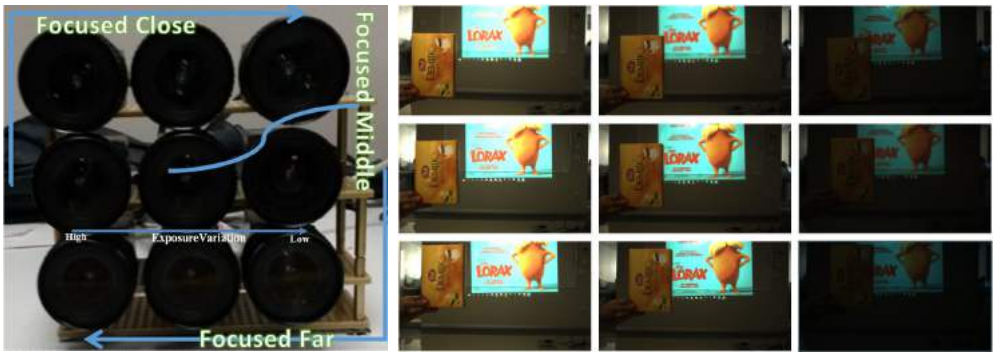
Focus stacking has also been the topic of interest for researchers over the past

and a large volume of papers have been published in this area [77],[78], [79], [80], [81], [82]. In order to generate an all in focus image from a set of differently exposed images, a spatial frequency based method is proposed in [77]. The method involves applying a sharpness measure on all the images to determine the pixels with larger information content in each image. Another class of methods for generating all in focus images are the image pyramid based methods. These methods [78] and [79], usually perform a multiscale decomposition of the image in order to identify the pixels or image regions with higher information content at different scales. Defocus modeling based methods [80], [81], [82] provide an alternative for all in focus image creation. These methods are based on an assumption of a known point spread function (PSF) model and to undo the effect of the PSF an inverse filter is applied resulting in an all in focus image. In wavelet transform based methods the wavelet decomposition of the series of focus images is performed which is followed by the fusion process in the wavelet domain and the wavelet coefficients are selected based on some criterion [83].

## 3.2 Camera array configuration

A 3 x 3 camera array is developed to simultaneously captures a total of 9 images with varying exposure, depth of field and largely overlapping field of view. In Figure 3.1, the diversification of the camera array is presented.

Allied machine vision cameras with a Ricoh 8.5mm lens are used in the entire camera array. Three different focus settings, focused close, focused middle and focused far are selected for the experiments. Focused close translates to 0.3 meters from the camera, focused middle translates to 0.5 meters from the camera, and focused far is anything above 0.5 meters. The top row of the camera array $C_{(1,j)}$ is focused close with the addition of the second row leftmost camera $C_{(2,1)}$. Camera $C_{(2,2)}$ is focused middle while the bottom row of the cameras $C_{(3,j)}$ along with camera $C_{(2,3)}$ is focused far. This configuration improves the registration of all the images with respect to the camera $C_{(2,2)}$ image because of the availability of images with a similar focus setting for both horizontal and vertical alignment.




(a) Camera array (b) Images captured

Figure 3.1: Camera array configured to capture multiple exposure images with varying depth of field. (a) Camera array marked with focus and exposure variation. (b) Images captured with the camera array shown in 3.1a.

The exposure of the cameras in a camera array varies along the rows. The left cameras $C_{(i,1)}$ have the highest exposure the middle ones $C_{(i,2)}$ have middle exposure while the right cameras $C_{(i,3)}$ have minimum exposure. To change the exposure across the row, the aperture opening is varied according to the camera position. To synchronize the acquisition of the frames all the cameras are connected with the external hardware trigger. PCI express slots and Vimba SDK are used to interface the camera array to the PC with Intel i5 @ 3.2 GHz processor and 16 GB RAM. The algorithm is implemented in MATLAB.

## 3.3 Focus stacking algorithm overview

The proposed algorithm has two major parts, focus stacking given images with different exposures and focus settings, and the dynamic range extension through focus stacked multiple exposure images. The focus stacking process is highlighted in Figure 3.2 and detailed below

All the input images $I_{(i,j)}$ are divided into three categories based on their exposures. These input images from different exposure categories are spatially registered to the reference image $I_{r(2,2)}$). After the spatial registration, the registered images$\tilde{I}_{(i,j)}$ are photometrically mapped to an image from a particular exposure category. Weight is estimated for each pixel in all the geometrically

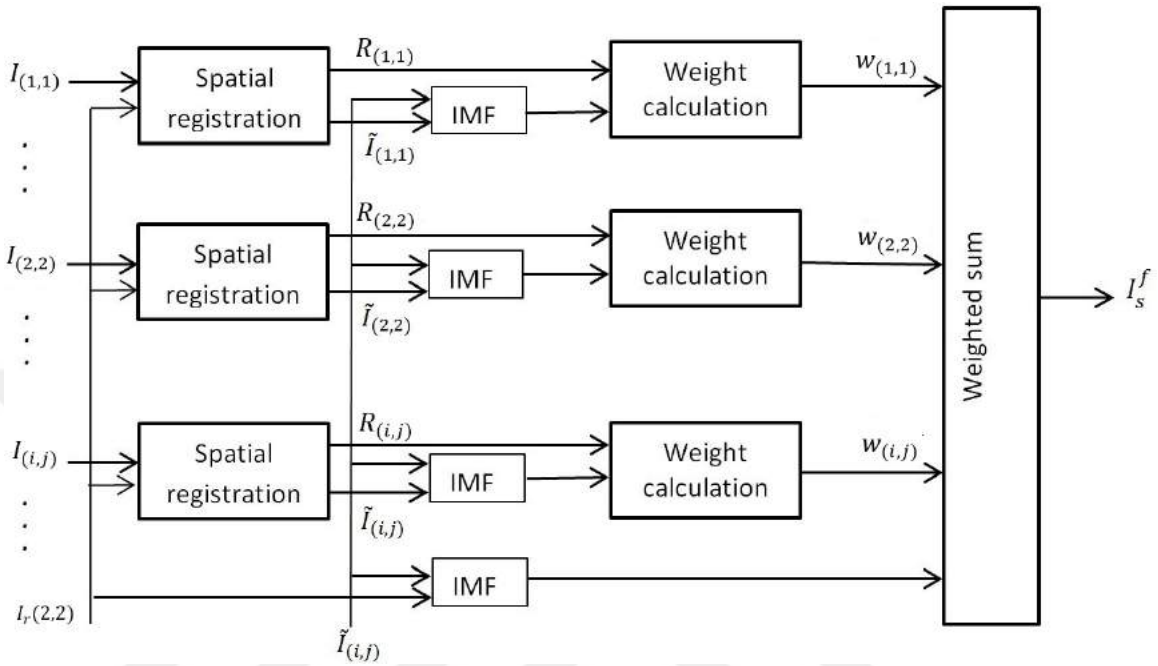


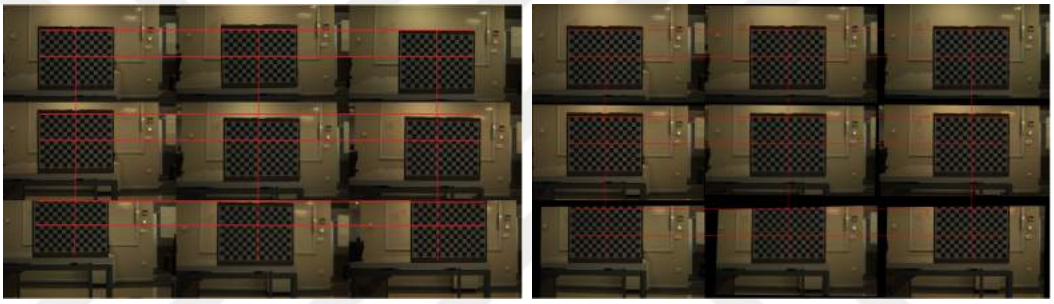

Figure 3.2: Proposed focus stacking algorithm flow. IMF is the intensity mapping function [13] and $i, j$ are the row and column index of the cameras in the camera array.

and photometrically mapped images. This weight is based on the two measures namely local sharpness and registration errors. The local sharpness measure of a pixel is large if it belongs to a focused area in an image and vice versa. However, a large registration error indicates that a pixel in the registered image is not correctly mapped to the corresponding pixel in the reference image $I_{r(2,2)}$. The weight is, therefore, the combination of local sharpness and the inverse of registration error. The pixels with the highest weight in all of the registered images is selected and combined with other selected pixels to generate a final focus stacked image $I_s^f$. The process is repeated for the remaining different exposure categories resulting in a total of three focus stacked images. The details of the steps involved in focus stacking are provided below.

## 3.4 Spatial registration

Combining all the different exposure and focus images into a single focus stacked HDR image requires all the images to be perfectly aligned. To have a large

overlapping area for all the images $I_{r(2,2)}$ is chosen as a reference image and all the remaining images are aligned with respect to the reference image. One simple way to achieve this task is to estimate the motion fields between the reference image $I_{r(2,2)}$ and the input images $I_{(i,j)}$. However, there is a difference in the focus of the images which may result in large registration error in the regions where there is a focus mismatch. To overcome this problem the camera array is rectified both horizontally and vertically so the resulting images form a regular grid as shown in Figure 3.3. Since the focus of the images is uniform across the



(a) Before rectification (b) After rectification

Figure 3.3: Camera array calibration for better spatial registration of the target images on the reference image.

top row as well as across the bottom row, the motion field across the x-axis is estimated between the middle image $I_{(1,2)}$ and the two adjacent images $I_{(1,1)}$ and $I_{(1,3)}$ for the top row and similarly for the bottom row between $I_{(3,2)}$ and the two corresponding adjacent images $I_{(3,1)}$ and $I_{(3,3)}$. Since the images are placed on a regular grid after rectification the same motion fields estimated for either the top or bottom row are used for aligning the middle row images $I_{(2,1)}$ and $I_{(2,3)}$ with $I_{r(2,2)}$. Similarly to vertically register the images, motion vector across the y-axis is estimated between image $I_{(1,1)}$ and $I_{(2,1)}$ and used for all the top row images. A motion vector between $I_{(2,3)}$ and $I_{(3,3)}$ is estimated and is used for all the bottom row images. In the non-overlapping part, some ghosting artifacts may appear in the final all in focus images which are handled during the fusion process and is explained in a section below.

Since there is a variation in the exposure of the cameras across a row, the optical flow algorithm cannot be directly applied to estimate the x-axis motion field as the constant brightness assumption does not hold anymore. In order to

fulfill the constant brightness requirement intensity mapping must be performed before optical flow estimation.

In our experiments histogram based intensity mapping function [13] is used because it is robust to such misalignments, apperas to be an appropirate choice. The motion fields are estimated using [84], because it has a robust data fidelity term and discontinuity preserving total variation regularization term. Once all the motion fields are estimated every input image is warped to the reference image $I_{r(2,2)}$ by using the corresponding motion fields. The warped images are now considered to be spatially registered with a reference image in the overlapping regions.

## 3.5 Extending depth of field

All the registered images $\tilde{I}_{(i,j)}$ are separated based on their exposures, resulting in three categories, high, middle, and low exposure with three images in each category. For each category an all in focus image $I_s^f$ is generated which has the depth of field larger than any of the individual images in that category. Since there is a variation of exposure, images from different exposure categories cannot be used as such. Therefore, all the registered images $\tilde{I}_{(i,j)}$ are photometrically mapped to a one of the image from a exposure category under consideration. Although all of the images within a category are expected to have the same exposure but there might be slight differences due to a minute change in the aperture opening and hence photometric registration is performed within a category as well.

As mentioned earlier the final all in focus image $I_s^f$ is generated by combining pixels from images captured with different focus settings. To choose these pixels a weight is assigned to each pixel based on its local sharpness and registration error. Local sharpness has been extensively used in common focus stacking algorithms to indicate the focus. The local sharpness is defined as follows

$$S(x) = \sum_{(y) \in W_h(x)} ||\nabla(\tilde{I}_{(i,j)}(y))||, \tag{3.1}$$

where, $\nabla(\tilde{I}_{(i,j)}(y))$ is a gradient vector at pixel $y$ obtained by applying Sobel filter, $||.||$ denotes gradient magnitude, and $W_h(x)$ is an $h$ x $h$ window across pixel $x$. Now that sharpness for each pixel is known the registration error is estimated. One may suggest the small threshold value at the absolute difference between the photometrically mapped registered image and reference image can be applied to eliminate the non-registered pixels, however, it should be noted that the images have different focuses and the absolute difference may be large even when the images are registered correctly. It is decided to use the motion vectors from spatial registration step between the reference $I_{r(2,2)}$ and input $I_{(i,j)}$ images. These motion vectors have significantly large values for pixels in the non overlapping or occluded regions which can be easily eliminated. It is noticeable that using flow vector based thresholding instead of simple absolute error based thresholding handles ghosting artifacts better in the final HDR image.

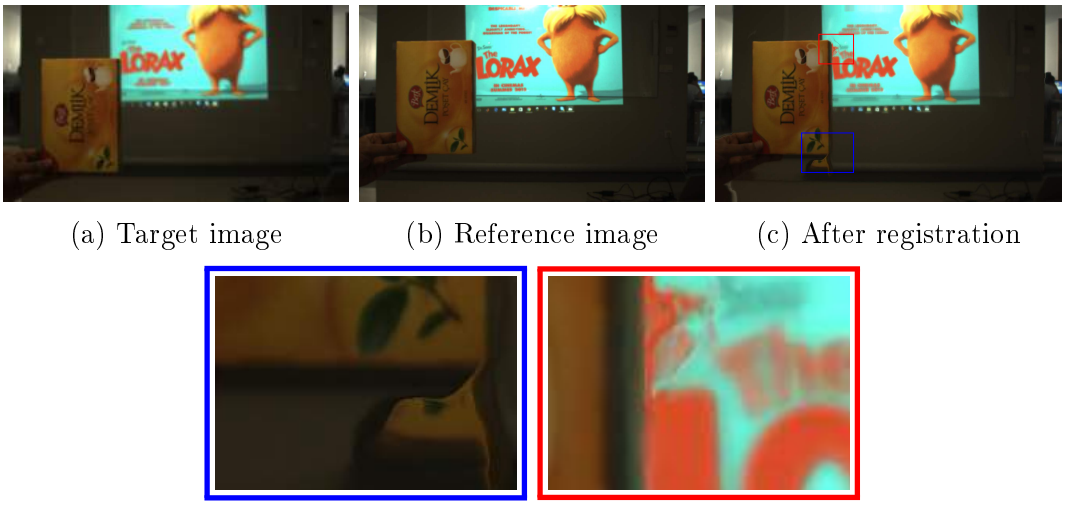
The thresholding process outputs either one or zero for every pixel in an image depending on its registration and saturation conditions resulting in a binary registration error map $R(i, j)$ for every image in a category. This registration error map is then combined with local sharpness map resulting in a final weight map $w(i, j)$. For common black pixels in all weight maps $w(i, j)$, the reference image $I_{r(2,2)}$ is photometrically mapped to an image in exposure category under consideration and the pixel is then selected from the intensity mapped reference image. Finally, the weighted sum of all the images performed to generate a final all in focus images $I_s^f$. The process is repeated three times to generate an all in focus image for each exposure category.

## 3.6 HDR radiance map estimation and tone mapping

Three all in focus images with different exposures can now be processed through any standard HDR image creation algorithm to generate an HDR image. In this work, a commercially available software Photomatix pro is used to estimate the radiance map. The software estimates the exposure time from the input images so there is no need to explicitly estimate the exposure time for each frame. To display the HDR image on an LDR display a standard tone compression technique provided by the software is used. The resulting image has a larger depth of field and high dynamic range than any of the individual images captured through the camera array.

## 3.7 Experimental results

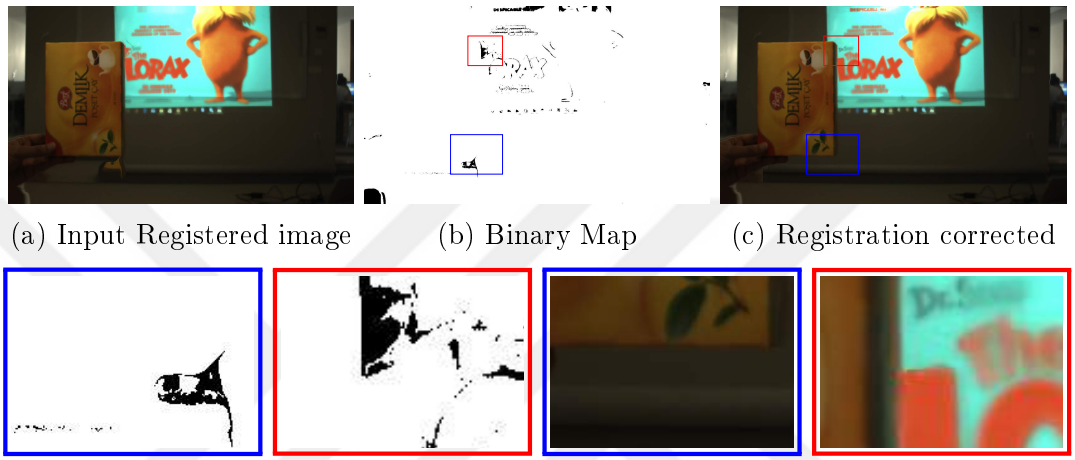
In this section, the experimental results of the proposed algorithm on a real dataset are provided. In Figure 3.4, the artifacts caused by the spatial registration of the target images on the reference image are shown.




(a) Target image (b) Reference image (c) After registration

Figure 3.4: Effects of the spatial registration process on one of the target image warped to reference image.

These artifacts are the result of estimating optical flow in the occluded regions. The motion vectors in these regions produce very large values and hence they are easily detected by the algorithm and eliminated as shown in the Figure 3.5.



(a) Input Registered image (b) Binary Map (c) Registration corrected

Figure 3.5: Minimizing artifacts caused by the spatial registration through proposed algorithm.

In Figure 3.6, a comparison of the final focus stacked high dynamic range image with individual multiple exposures, multiple focus images is presented to demonstrate the extended depth of field and dynamic range.

As mentioned earlier three different focus setting ranging from 0.3 meters away from the camera to 0.5 meters and beyond 0.5 meters are used to capture images and hence the all in focus image has a depth of field extended by 20 cm from the image captured with the camera that has the largest depth of field in the camera array. To provide an idea of the dynamic range extension by the camera array the histograms of, high, mid, and low exposure images with the HDR image histogram are compared in Figure 3.7. The noise is relatively higher in darker regions, so the effective dynamic range is considered from zero onwards.

Estimating motion flow based registration error instead of the traditional absolute difference method a significant reduction is achieved in the ghosting artifacts. In Figure 3.8, a comparison the proposed method with the state of the art HDR image creation methods is presented. All the methods involve deghosting step and the proposed method show significant improvement over these methods.

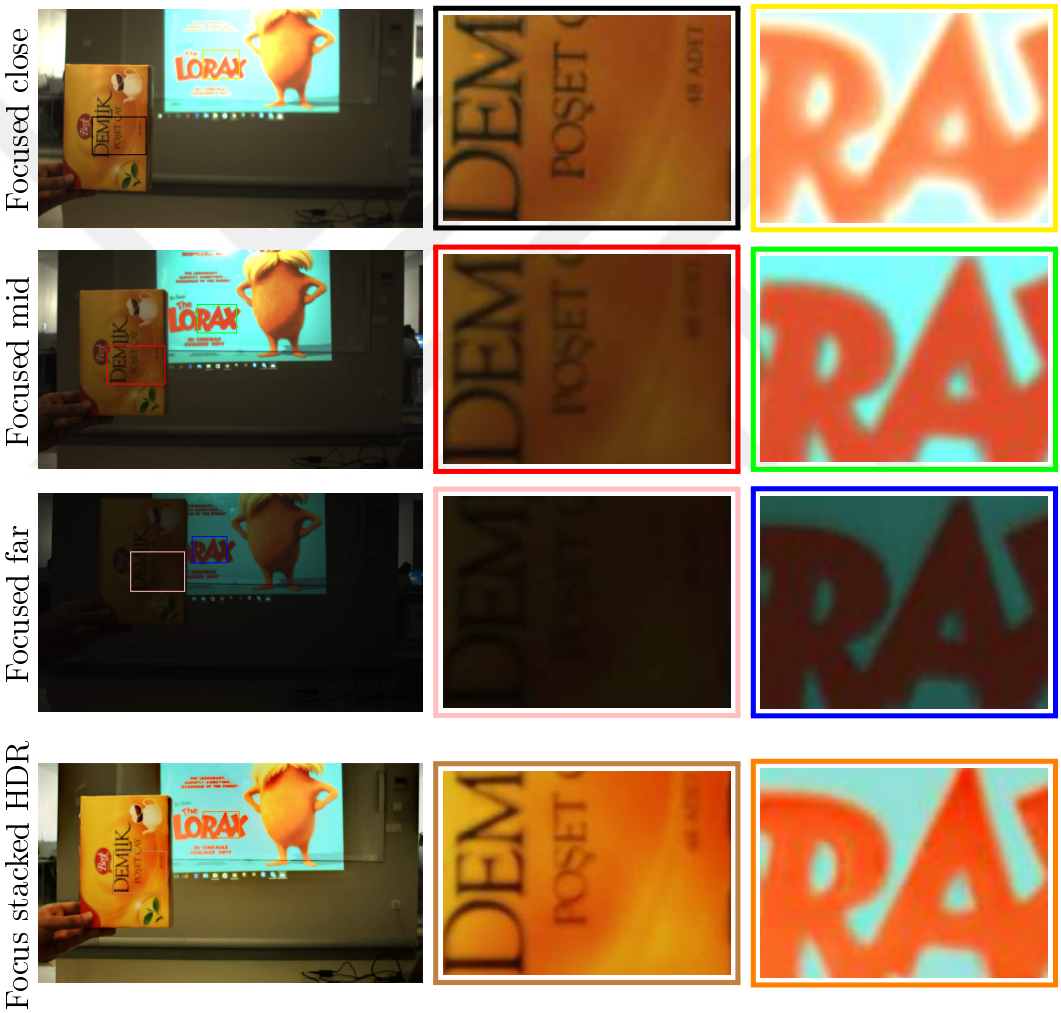
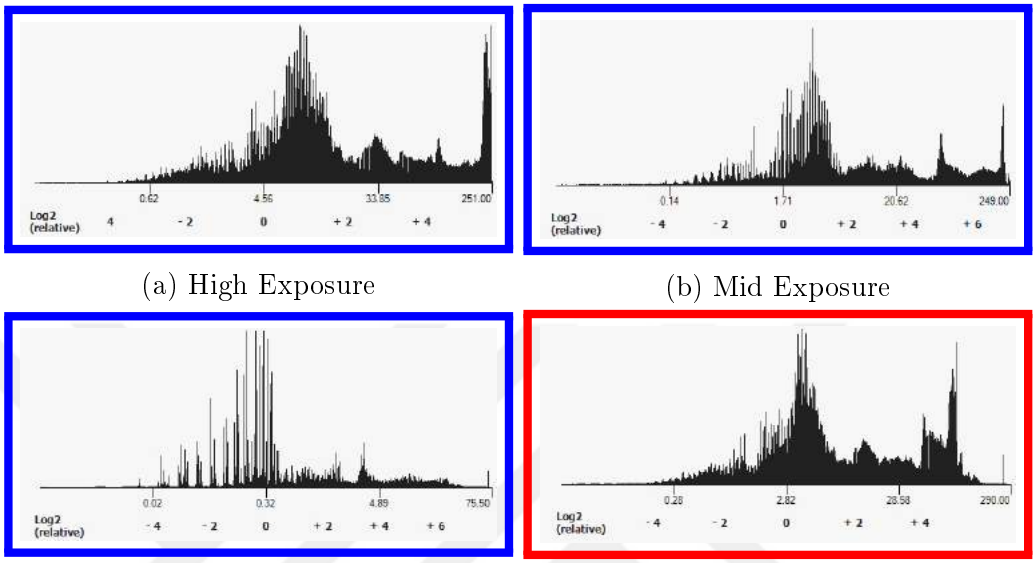



Figure 3.6: Depth of field and Dynamic range comparison of the individual images captured by the camera array with the focused stacked HDR image generated by the proposed algorithm.




(a) High Exposure (b) Mid Exposure

(c) LOW Exposure (d) HDR

Figure 3.7: Comparison of the histogram of multiple-exposure LDR images with HDR image.

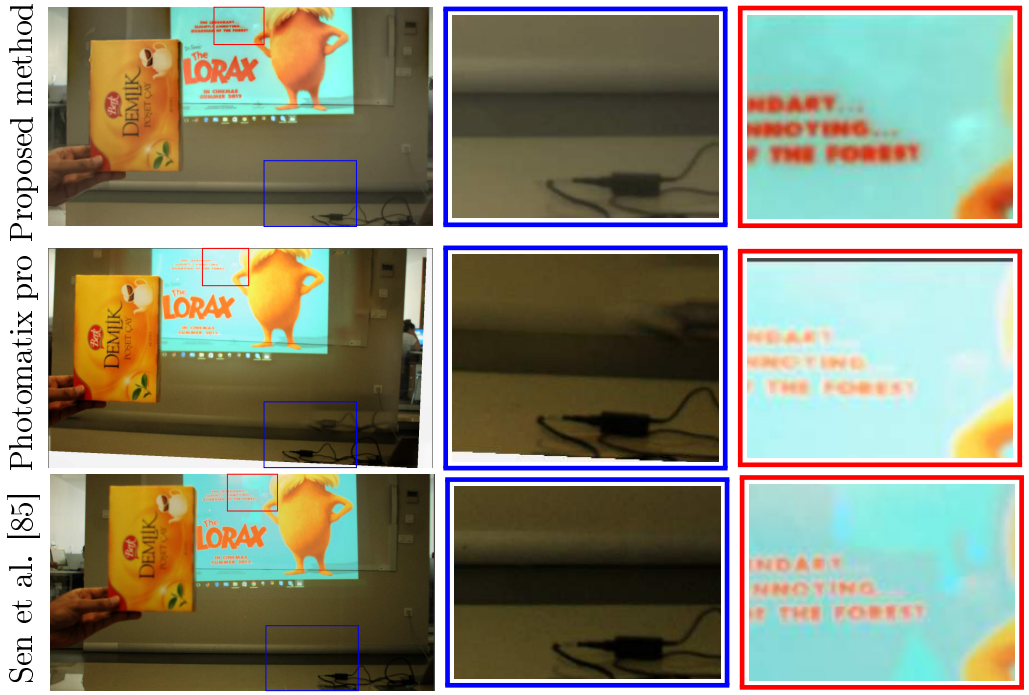



Figure 3.8: Comparison of the proposed method with state of the art HDR creation techniques that include deghosting algorithm.

# Chapter 4

# Hybrid Light Field Imaging for Improved Spatial Resolution and Depth Range

Light field imaging involves capturing both angular and spatial distribution of light; it enables new capabilities, such as post-capture digital refocusing, camera aperture adjustment, perspective shift, and depth estimation. Micro-lens array (MLA) based light field cameras provide a cost-effective approach to light field imaging. There are two main limitations of MLA-based light field cameras: low spatial resolution and narrow baseline. While low spatial resolution limits the general purpose use and applicability of light field cameras, narrow baseline limits the depth estimation range and accuracy. In this chapter, a hybrid stereo imaging system that includes a light field camera and a regular camera is presented. The hybrid system addresses both spatial resolution and narrow baseline issues of the MLA-based light field cameras while preserving light field imaging capabilities.

## 4.1 Related work

*On low spatial resolution:* There are various methods proposed to address the low spatial resolution issue in MLA-based light field cameras. One main approach is to apply super-resolution image restoration to light field sub-aperture images. Super-resolution in a Bayesian framework is commonly used, for example, in [86] with Lambertian and textural priors, in [87] with a Gaussian mixture model, and in [88] with a variational formulation. Learning-based methods are adopted as well, including dictionary-based learning [15] and deep convolutional neural networks [89, 90]. In addition to spatial domain super-resolution restoration, Fourier-domain techniques [91, 92] and wave optics based 3D deconvolution methods [93, 94, 95, 96] have also been utilized.

Alternative to the standard MLA-based light field camera design [6], where the MLA is placed at the image plane of the main lens and the sensor is placed at the focal length of the lenslets, there is another design approach where the MLA is placed to relay image from the intermediate image plane of the main lens to the sensor [22]. This design is known as "focused plenoptic camera." As in the case of the standard light field camera approach, super-resolution restoration for focused plenoptic cameras is also possible [97].

All single-sensor light field imaging systems are fundamentally limited by the spatial-angular resolution trade-off, and the above-mentioned restoration methods have performance limitations in addition to the computational costs. Another approach for improving spatial resolution is to use a hybrid two camera system, including a light field camera and a high-resolution camera, and merge the images to improve spatial resolution [7, 98, 99]. Dictionary-learning based techniques are adopted [7, 98] in this problem as well: High resolution image patches from the regular camera are extracted and stored as a high resolution patch dictionary. These high resolution patches are downsampled; and from the downsampled pathes, low resolution features are extracted to form a low resolution patch dictionary. During super-resolution reconstruction, a low resolution image patch is enhanced through determining (based on feature matching) and

using the corresponding high resolution patches in the dictionary. In [99], high resolution image is decomposed with complex steerable pyramid filters; the depth map from the light field is upsampled using joint bilateral upsampling; perspective shift amounts are estimated from the upsampled depth map, and these shift amounts are used to modify the phase of the decomposed high resolution image; with the modified phases, pyramid reconstruction is applied to obtain high resolution light field.

*On narrow baseline:* One of the most important features of light field cameras is the ability to estimate depth. However, it is known that depth accuracy and range is limited in MLA-based light field cameras due to narrow baseline. The relation between baseline and depth estimation accuracy in a stereo system has been studied in [100]. In a stereo system with focal length $f$ and baseline $b$, the depth $z$ of a point with disparity $d$ is obtained through triangulation as $z = fb/d$. With a disparity estimation error of $\epsilon_d$, the depth estimation error $\epsilon_z$ becomes [100]:

$$\epsilon_z = \frac{fb}{d} - \frac{fb}{d + \epsilon_d} = \frac{d^2 \epsilon_d}{f * b + d\epsilon_d} \approx \frac{z^2}{fb} \epsilon_d, \tag{4.1}$$

which indicates that the depth estimation error is inversely proportional with the baseline and increases quadratically with depth. The disparity error $\epsilon_d$ is typically set to 1, and the depth estimation error $\epsilon_z$ as a function depth can be calculated. It is also possible to set an error bound on $\epsilon_z$ and derive the maximum depth range from the above equation.

For an MLA-based light field camera, the maximum baseline is less than the size of of the main lens aperture, making depth estimation very challenging. There are methods specifically proposed for depth estimation in MLA-based light field cameras. For example, in [101], the problem is formulated as a constrained labeling problem on epipolar plane images in a variational framework. In [28], ray geometry of 3D line segments is imposed as constraints on light field triangulation and stereo matching. In [30], defocus and shading cues are used to improve the disparity estimation accuracy.

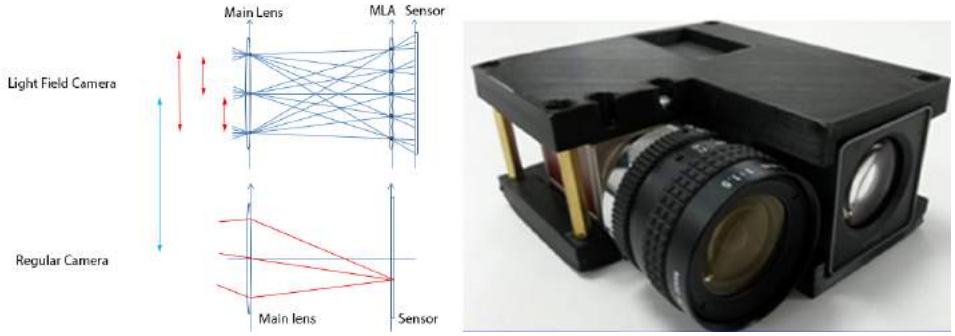



Figure 4.1: Hybrid imaging system including a regular and a light field camera. The maximum baseline of the light field camera is limited by the camera main lens aperture, and is much less (about an order of magnitude) than the baseline (about 4cm) between the light field and the regular camera.

(a) (b)

Figure 4.2: (a) Raw light field. (b) Decoded sub-aperture images.

## 4.2 Hybrid stereo imaging

The hybrid stereo imaging system consists of a regular camera and a light field camera as shown in Figure 4.1.

The system has two advantages over a single light field camera: (i) The high-resolution image produced by the regular camera is used to improve the spatial resolution of each sub-aperture image extracted from the light field camera. That is, a light field with enhanced spatial resolution is obtained. (ii) The large baseline between the regular camera and the light field camera results in a wider range and more accurate depth estimation capability, compared to a single light field camera.

(a) (b) (c)

Figure 4.3: (a) Regular camera image. (b) Regular camera image after photometric registration. (c) One of the bicubically resized Lytro sub-aperture image.

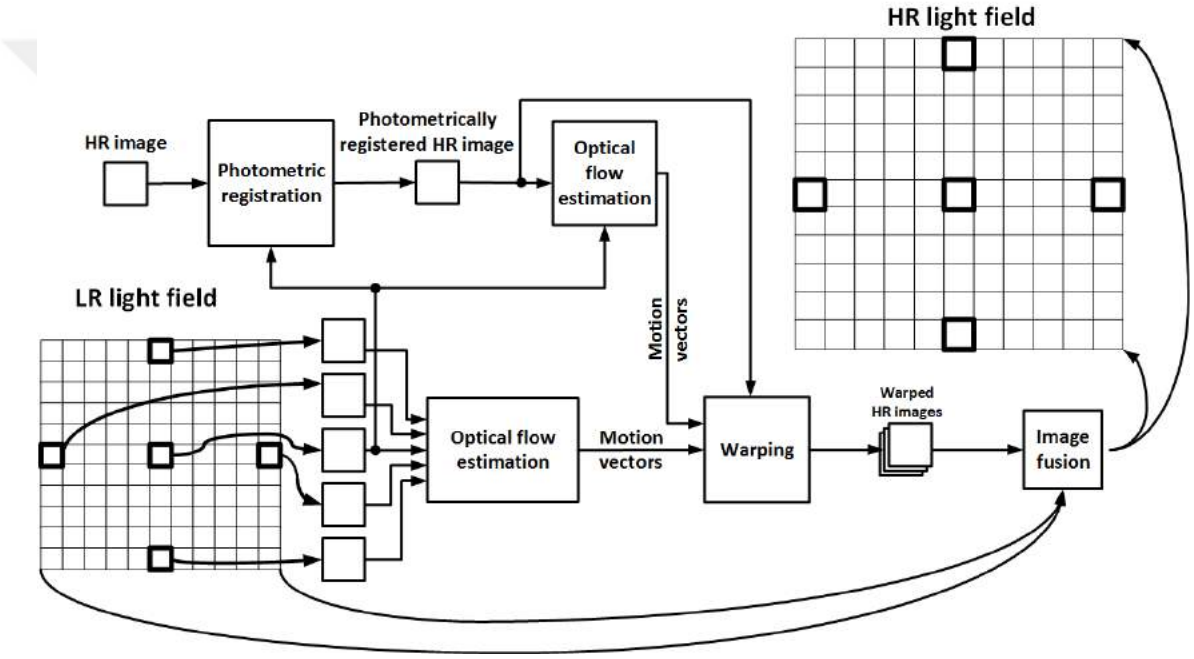



Figure 4.4: Illustration of the resolution enhancement process.

### 4.2.1 Prototype system and initial light field data processing

The prototype system includes a first-generation Lytro camera and a regular camera (AVT Mako G095C). The light field is decoded using [14] to obtain 11x11 sub-aperture images, each with size 380x380. The regular camera has a spatial resolution of 1200x780 pixels. The imaging system is first calibrated: The regular camera image and the light field middle sub-aperture image is calibrated (utilizing the Matlab Stereo Calibration Toolbox) to determine the overlapping regions between the images and rectify the regular camera image. The regular image is then photometrically mapped to the color space of the light field sub-aperture

images using the histogram-based intensity matching function technique [102].

A raw light field data and the extracted sub-aperture images are shown in Figure 4.2. In Figure 4.3, the rectified regular camera image is shown along with a light field sub-aperture image.

### 4.2.2 Improving spatial resolution

An illustration of the resolution enhancement process is given in Figure 4.4. Each low-resolution (LR) light field sub-aperture image is bicubically interpolated to match the size of the high-resolution (HR) regular camera image. The optical flow between the HR image and the light field middle sub-aperture image and the optical flow between the light field middle sub-aperture and every other sub-aperture images are estimated. (The optical flow estimation algorithm presented in [103] is used in all experiments.) Combining these optical flow estimates, motion vectors between the HR image and each light field sub-aperture image are obtained. The HR image is warped onto each light field sub-aperture image and fused to produce a high-resolution version of each sub-aperture image. As a result, a high-resolution light field is obtained.

The problem of image fusion for resolution enhancement has been well studied in the literature, with applications in satellite imaging for pan-sharpening, digital camera pipelines for demosaicking, and in computational photography for multi-focus stacking [104]. Two basic methods for image fusion are used: (i) a wavelet-based approach [105], available in Matlab as function *wfuseimg*, which essentially replaces the detail subbands of low-resolution image with the detail subbands of the high resolution image, and (ii) alpha blending, also available in Matlab as function *imfuse*, which simply takes the weighted average of input images.

*Speeding up the registration process:* The speed of the registration process is further increased by using the fact that light field sub-aperture images are captured on a regular grid. Instead of estimating the optical flow between the middle sub-aperture image and each of the remaining sub-aperture images, the

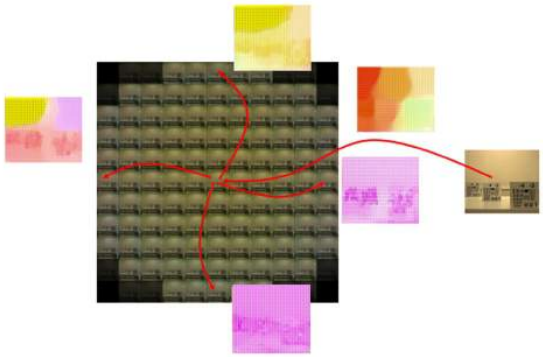


Figure 4.5: Speeding up the optical flow estimation process.

optical flow is estimated between the middle and the leftmost, rightmost, topmost, and bottommost sub-aperture images as shown in Figure 4.5. The estimated motion vectors are interpolated for the rest of the sub-aperture images according to their relative positions within the light field. As a result, four within-light-field-camera optical flow estimation (instead of 120) and one between-cameras optical flow estimation are obtained. In Figure 4.6, the difference between the regular camera image and light field sub-aperture images before and after registration is shown. The optical flow within the light field is estimated as described above. The after-registration result shows that the registration process works well. (Note that the residuals for the sub-aperture images in the aperture corners are large because of the fact that the original sub-aperture images in the corners are too dark due to vignetting.)

## 4.3 Experimental results

In this section, experimental results on resolution enhancement and depth estimation are presented. All implementations are done with Matlab, running on an Intel i5 PC with 12GB RAM. For the resolution enhancement process, given in Figure 4.4, the processing time of an entire Lytro light field is about 70 seconds, in which the optical flow estimation per image pair is about 11 seconds. In Figure 4.7, a comparison of light field sub-aperture image with its resolution-enhanced version is presented. Both the alpha blending and wavelet-based image

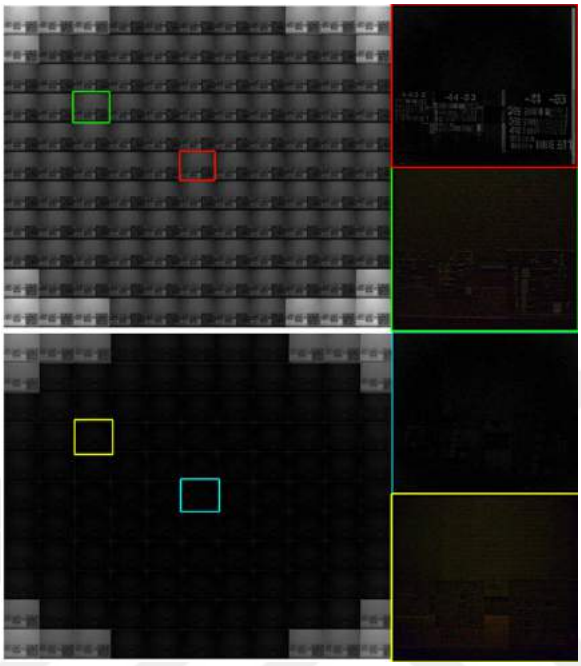
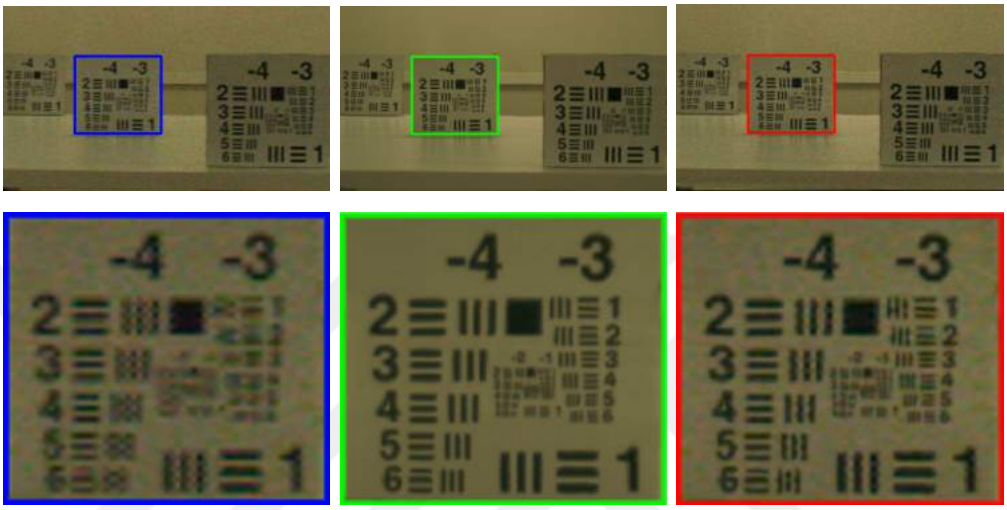


Figure 4.6: (Top) Residual between the regular camera image and light field sub-aperture images before warping. Two sub-aperture images are highlighted. (Bottom) Residual between the regular camera image and light field sub-aperture images after warping.

fusion processes produce good results in terms of resolution enhancement. Alpha blending suppresses the low-spatial-frequency color noise better than the wavelet-based approach; this is expected because the wavelet-based approach preserves the low-frequency content of the light field images, which have more noise compared to the image obtained from the regular camera. Alpha blending, on the other hand, simply averages two images, reducing the overall noise in all parts of the final image. (In all experiments, the weights of the HR image and light field sub-aperture images are 0.55 and 0.45, respectively, giving slightly more weight to the HR image in alpha blending.)

In [7] it has demonstrated that hybrid imaging approach has better performance in terms of image quality as compared to single sensor based restoration techniques but their specific algorithm is computationally expensive. With the proposed stereo setup and optical flow based registration the computational time is greatly reduced without compromising the quality of the high resolution light



(a) (b) (c)

Figure 4.7: Resolution enhancement of light field sub-aperture images. (a) One of the bicubically resized Lytro sub-aperture image. (b) Resolution-enhanced sub-aperture image using alpha blending. (c) Resolution-enhanced sub-aperture image using wavelet-based fusion.

(a) (b) (c) (d)

Figure 4.8: Comparison with the existing light field resolution enhancement algorithms. (a) Dansereau et al. [14] (b) Cho et al. [15] (c) Boominathan et al. [7] (d) Proposed.

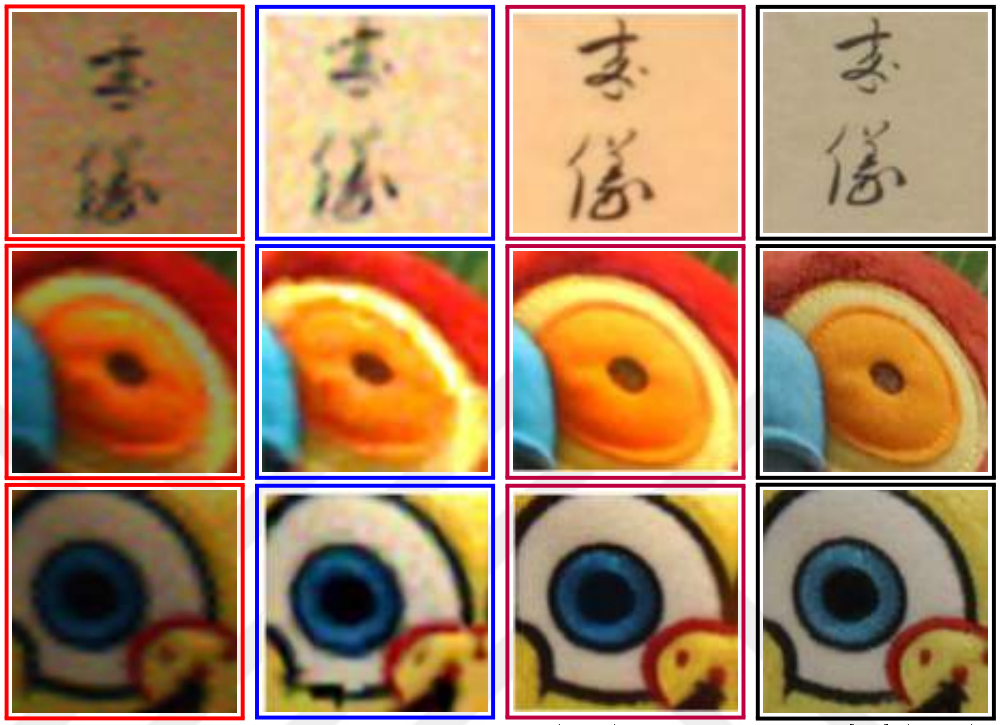
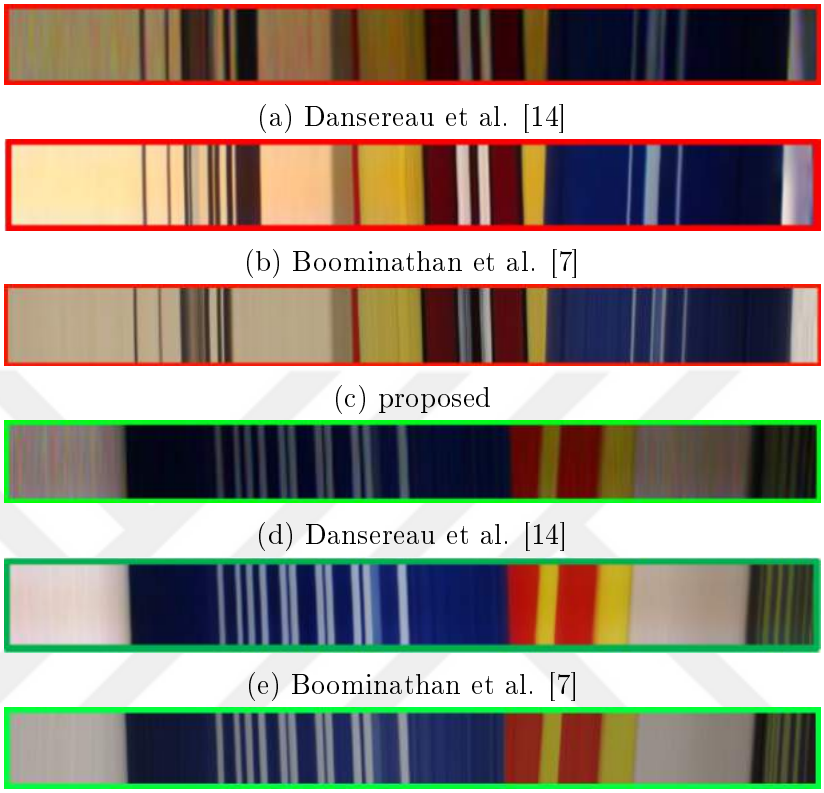


Figure 4.9: Zoomed in regions from Figure 4.8. (Red) Dansereau et al. [14] (Blue) Cho et al. [15] (Purple) Boominathan et al. [7] (Black) Proposed.

field. In figure 4.8 - 4.10 a comparison the image quality of the resolution enhanced light field with the existing methods is presented.

*Refocusing:* One of the key features of light field imaging is post-capture digital refocusing through a simple shift-and-sum procedure [4]. In Figure 4.11, refocusing at different distances with Lytro light field images and the resolution-enhanced light field sub-aperture images is shown. It can be clearly seen that sharper refocusing is attained through the proposed algorithm as compared to the original Lytro images. In Figure 4.12, refocusing examples from another data set captured by the proposed imaging system is presented. Again, the resolution-enhanced light field result in higher resolution refocusing compared to the Lytro light field.

*Addressing Occlusion:* The registration based on optical flow leads to holes on the occlusion boundaries. Hence the fused image loses high frequency signal in boundary regions. these high frequency signals are retained by first detecting the holes using a binary hole mask. In the mask those pixels which are present



(a) Dansereau et al. [14]

(b) Boominathan et al. [7]

(c) proposed

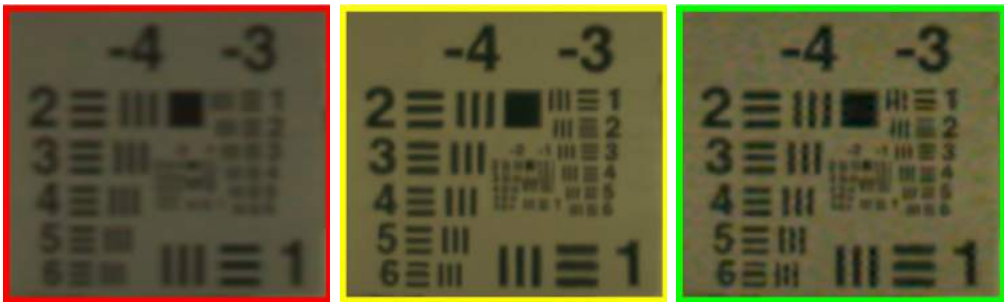
(d) Dansereau et al. [14]

(e) Boominathan et al. [7]

(f) Proposed

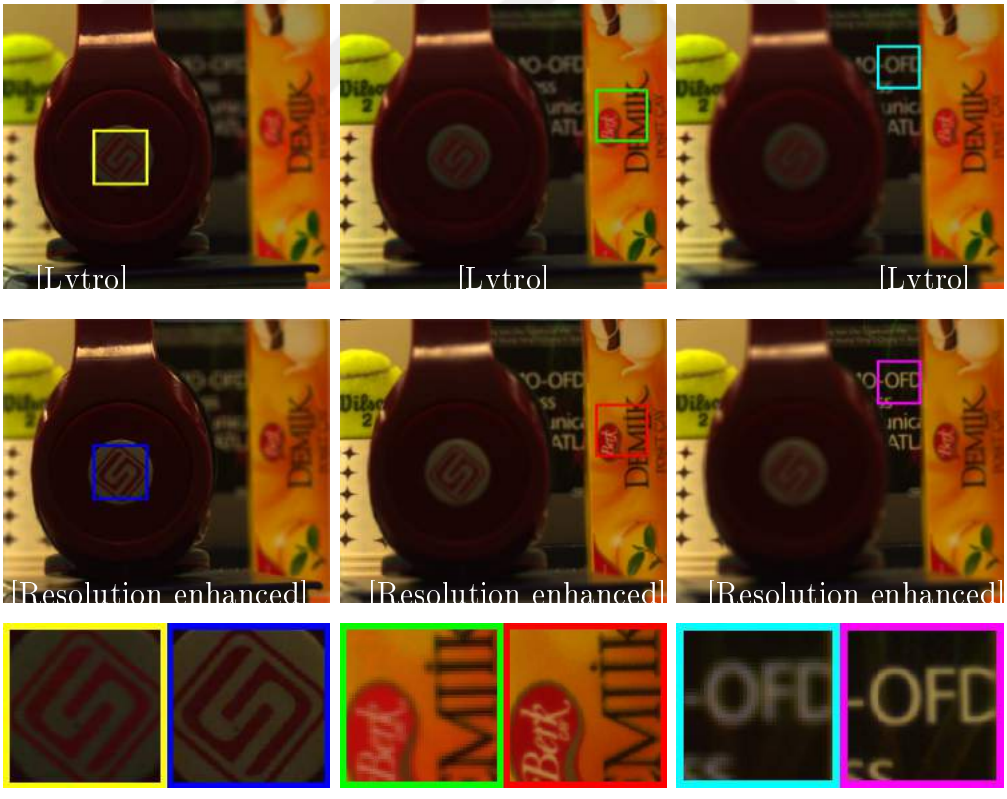
Figure 4.10: Comparison of the EPI images with the existing light field super resolution techniques

in re-sized sub-aperture image and are missing in the warped high resolution image are set to one while the non-missing pixels are set to zero. Hole filling is then performed by taking the occluded regions in the fused image from the re-sized sub-aperture image. Since the holes are only present at the boundaries any significant degradation in the overall quality of the fused image is not noticed and the high frequency content is preserved as shown in Figures 4.13 and 4.14. *Improving depth range and accuracy:* To demonstrate the increased depth range and improved depth estimation accuracy of proposed hybrid imaging system, an experimental setup is we devised, where target objects (i.e, "Lego blocks") are placed in the scene starting from 40cm away from the imaging system. In Figure 4.15, the leftmost and rightmost light field sub-aperture images as well as the regular camera image is shown, in addition to the disparity maps, which are estimated using [103]. The dynamic range of the disparity map for the hybrid system (between the middle sub-aperture image and the regular camera image) is




(a) (b) (c)

Figure 4.11: Post-capture digital refocusing using shift-and-sum technique. (a) Lytro light field refocusing. (b) Resolution-enhanced (using alpha blending) light field refocusing. (c) Resolution-enhanced (using wavelet-based fusion) light field refocusing. (Bottom row) Zoomed-in regions from middle depth focusing.




(a) (b) (c)

Figure 4.12: Post-capture digital refocusing using the shift-and-sum technique. (Top row) Lytro light field refocusing. (Middle row) Resolution-enhanced light field (using alpha blending) refocusing. (Bottom row) Comparison of zoomed-in regions. (a) Close-depth focus. (b) Middle-depth focus. (c) Far-depth focus.

Figure 4.13: Hole filling after warping high resolution image on light field.

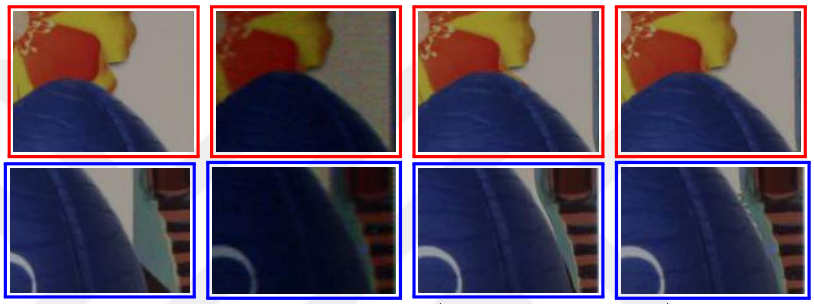


Figure 4.14: Zoomed in region of figure 4.13. (Left most column) high resolution. (Left column) low resolution. (Right column) Resolution enhanced. (Right most column) After hole filling.

about eight times larger than that of the light field camera (between the leftmost and rightmost sub-aperture images). Comparing the ranges of disparity maps and the separation of different objects from different depths, it can seen that the hybrid system improves the depth estimation accuracy. In Figure 4.15(f), the disparities for the target object positions are plotted. For the light field camera, the disparity difference from one depth to another becomes too small beyond 100cm, making it difficult to distinguish between different depths, and the disparities eventually become sub-pixel beyond 200cm. On the other hand, for the hybrid system, the disparities are large and distinguishable in the same range.

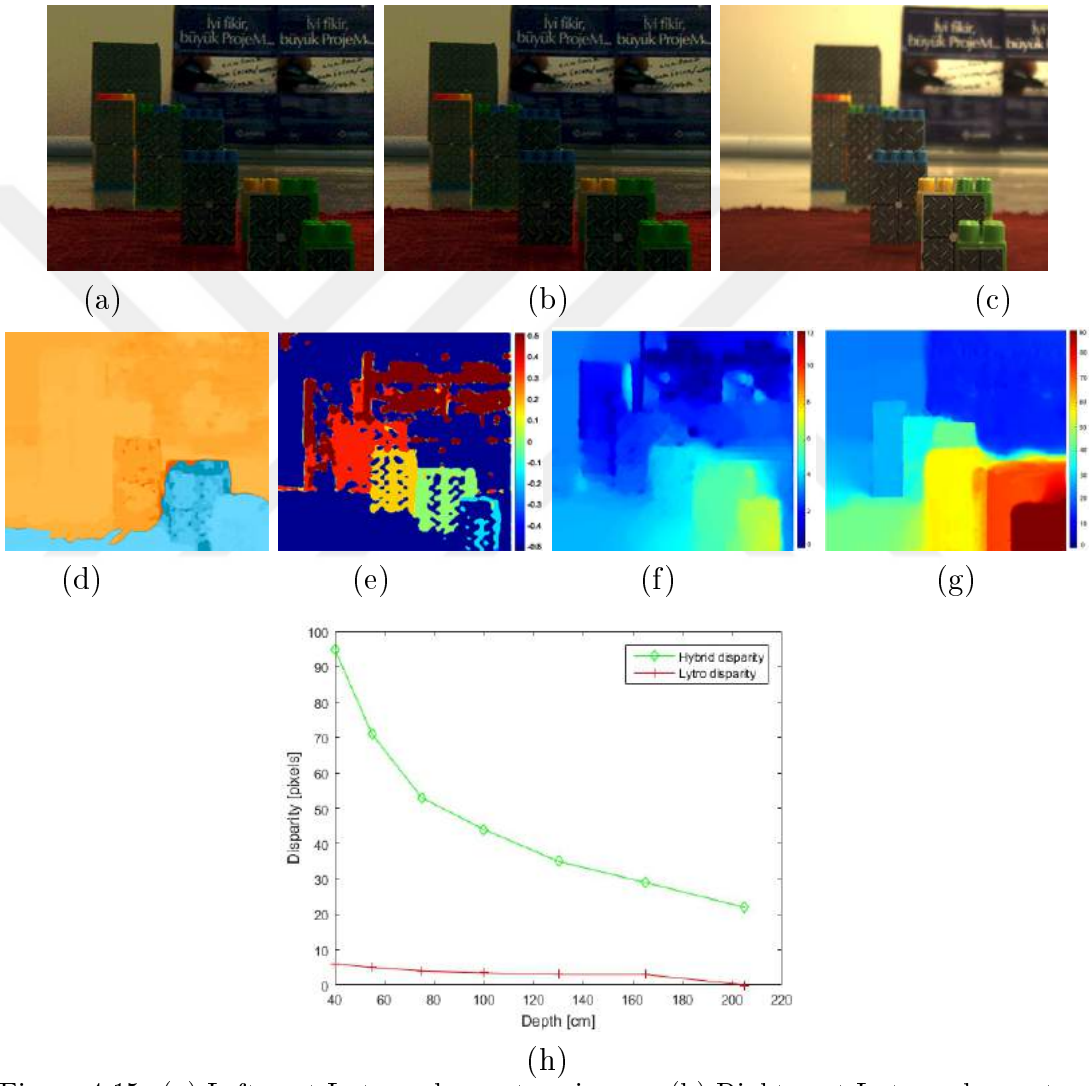



Figure 4.15: (a) Leftmost Lytro sub-aperture image. (b) Rightmost Lytro sub-aperture image. (c) Regular camera image (before photometric registration). (d) Disparity map by Lytro desktop software. (e) Slope based EPI disparity map [16]. (f) Disparity map between the leftmost and rightmost Lytro sub-aperture images. (g) Disparity map between the middle Lytro sub-aperture image and the regular camera image. (h) Disparities of the target object centers.

# Chapter 5

# Deconvolution Based Light Field Extraction from A Single Image Capture

In this chapter, a method to extract light field using a conventional camera from a single image capture is presented. The method involves an offline calibration process, where point spread functions, relating different perspective images captured with a narrow aperture to a central image captured with a wide aperture, are estimated for different depths. During application, light field perspective images are recovered by deconvolving the input image with the set of point spread functions that were estimated in the offline calibration process.

## 5.1 Related work

The proposed light field extraction technique requires depth estimation from a single blurry image. Estimating depth from a single defocused image has been studied extensively in[106, 107, 108, 109, 110, 111]. Single image captured with circular aperture usually assumes some prior information about either the PSF

[106], texture [107] or color information [108]. Even though some good results have been attained using DFD techniques with circular apertures, there are some inherent limitations.

Coded aperture photography has been used for both, improving out of focus deblurring [112, 113, 114] and better depth estimation [109, 110, 111]. Techniques that use coded aperture for depth estimation can be categorized into two types: multiple images/shots coded aperture methods and single image coded aperture methods.

In [115] coded aperture pair was optimized for optimal DFD performance by increasing relative defocus which depends on the difference in amplitude and phase spectra. In [116] a stereo camera system with coded apertures is designed to combine two types of depth cues i.e defocus and disparity. Both defocus and correspondence cues are also combined through Markov random field global optimization process in [117] using light-field data.

Also in single image-based methods, the advantages of the coded aperture designs are well utilized. In [109] the aperture shape is optimized by using Gaussian prior on image derivates. With this model, a likelihood for a blurry image to belong to a certain depth level can be derived. Therefore the Mask that has maximum Kullback Leibler divergence for this likelihood has better depth discriminability. In [110] a parametric maximization problem is defined to find a suitable mask pattern that differentiates between the images blurred with different blur scales. Two different objective functions are defined in [111] to find out a mask pattern that minimizes deblurring error with correct PSF but also maximizes deblurring error with incorrect PSFs. The proposed depth discrimination criterion is inspired from [109], however, a conventional aperture is utilized to capture the blurry image.

## 5.2 Offline PSF estimation for different perspectives and depths

The offline calibration process is illustrated in Figure 5.1. A planar object is placed at a distance $d_j$ from the camera. The aperture of the camera is closed to a narrow opening and the camera is moved to a perspective position $p_i$. The image corresponding to that specific perspective and depth is then captured and is denoted as $I_{i,j}(x, y)$, where the subscripts $(i, j)$ indicates the perspective position $p_i$ and the object depth $d_j$, and $(x, y)$ indicates a pixel coordinate. In addition to the narrow aperture perspective images, an image with a wide aperture, denoted as $B_j(x, y)$, is also taken for the same depth $d_j$. The process is repeated for all depths and perspectives. The perspective positions are chosen on a regular grid within the wide aperture opening.

The wide aperture image $B_j(x, y)$ can be modeled as the superposition of narrow aperture images taken from different perspectives within the aperture. Other than the occluded regions, the image $B_j(x, y)$ can be written as the convolution of the narrow aperture image $I_{i,j}(x, y)$ with a PSF $k_{i,j}(x, y)$ that incorporates the difference between the aperture sizes for $B_j(x, y)$ and $I_{i,j}(x, y)$: $B_j(x, y) = k_{i,j}(x, y) * I_{i,j}(x, y)$. The convolution equation would result in a set of linear equations when written for all $(x, y)$. These equations can be written as $\mathbf{B}_j = \mathbf{I}_{i,j}\mathbf{k}_{i,j}$, where $\mathbf{B}_j$ and $\mathbf{k}_{i,j}$ are the vectorized forms of the wide aperture image and the PSF, and $\mathbf{I}_{i,j}$ matrix is constructed from $I_{i,j}(x, y)$. The system using the least squares estimation technique to obtain the PSF for a specific perspective and depth. The PSF estimation process is repeated for all depths and perspectives.

The prototype system is shown in Figure 5.2; it includes two motorized translation stages (Thorlabs NRT150) and a regular camera with a 1024 x 1280 CMOS sensor and a 35mm lens. During the calibration, the planar target object is moved from 30cm away from the camera upto 2 meters with steps of 10cm. For each depth, the camera is moved with a step size of 0.1mm as shown in the figure to capture the perspective images; in addition, the camera is moved to the central

position to capture the wide aperture image.

## 5.3 Deconvolution based light field extraction

Once the PSFs are estimated, the camera is taken off from the translation stage, the aperutre is set wide, and it is ready for light field acquisition.

### 5.3.1 PSF scale identification

Light field extraction from a single wide aperture image capture consists of several steps. First, for each perspective, a depth (PSF scale) map is obtained. The depth map is then utilized to fuse multiple deconvolved images to construct a perspective image. The process is repeated for all perspectives.

The PSF scale identification algorithm is adopted from [109], which exploits the reconstruction error when there is scale and blur mismatch. When an image of a natural scene with multiple depths is deconvolved with the PSF of a particular scale, the corresponding depth regions in the image become sharp while severe ringing artifacts appear in the rest of the image because these areas cannot be explained by that scale of the PSF. These ringings are dense and have gradients with magnitude significantly larger than that of natural sparsely distributed data; the difference can be used as a cue for kernel scale identification [109, 118].

First, $L_{i,j}(x, y)$ are estimated by minimizing the following cost function.

$$|C_{k_{i,j}(x,y)}L_{i,j}(x,y) - B_j(x,y)| + \sum_{xy} \rho(L_{i,j}(x,y)... \\ - L_{i,j}(x+1,y)) + \rho(L_{i,j}(x,y) - L_{i,j}(x,y+1)), \quad (5.1)$$

where $\rho$ is a heavy-tailed function and $|C_{k_{i,j}(x,y)}L_{i,j}(x,y) - B_j(x,y)|$ is the reconstruction error and C represents the convolution between $k_{i,j}(x, y)$ and $L_{i,j}(x, y)$.

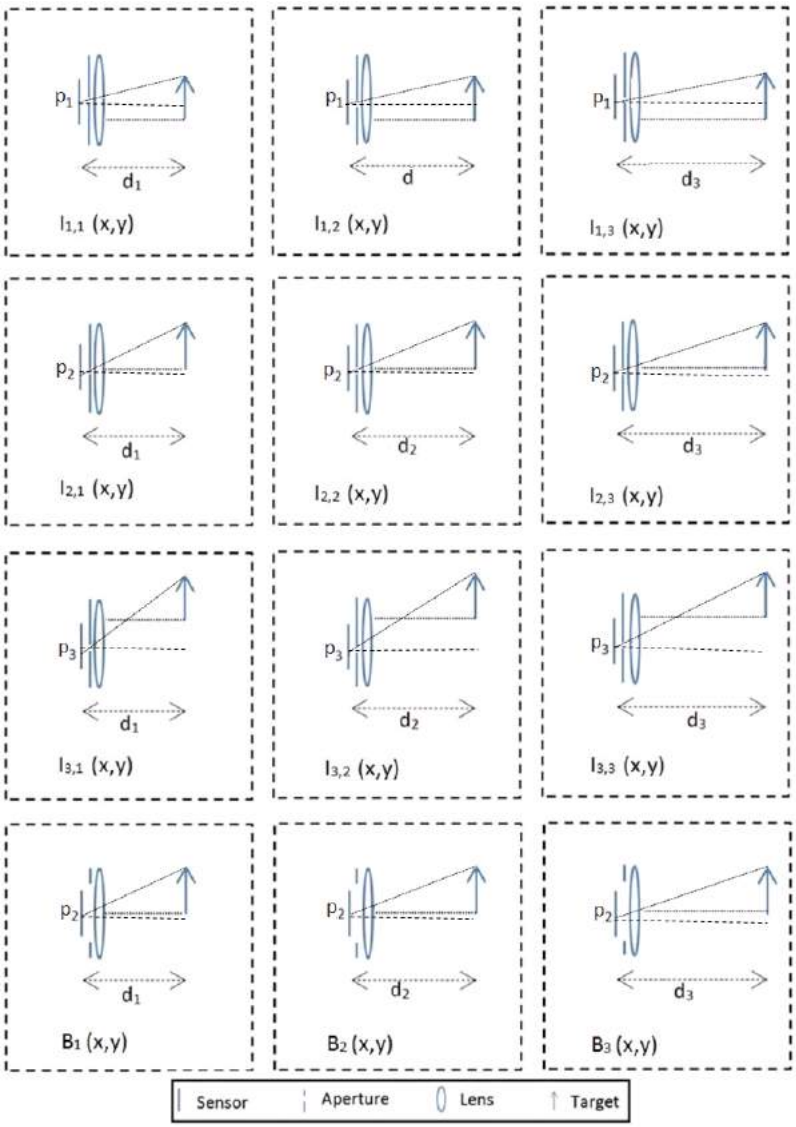



Figure 5.1: Illustration of the offline calibration process. For each depth $d_j$, narrow aperture images $I_{i,j}(x, y)$ from different perspective locations $p_i$ and a single wide aperture image $B_j(x, y)$ are taken. The process is repeated for different depths.

Figure 5.2: Setup for the offline calibration process. (Left) Motorized translation stages and the camera. (Right Top) Narrow aperture opening for perspective images. The camera is moved as illustrated on lens for different perspective images. (Right Bottom) Wide aperture opening for the central image.

Once the deblurred images are obtained, then for every pixel in the blurry image $B(x, y)$ the reconstruction error with all the sample PSFs $k_{i,j}(x, y)$ is estimated as follows.

$$\hat{E}_{i,j}(x, y) = \sum_{x,y \in W} |B(x, y) - k_{i,j}(x, y) \otimes L_{i,j}(x, y)|^2 ... + \lambda |\nabla L_{i,j}(x, y)|. \quad (5.2)$$

where $L_{i,j}(x, y)$ is the set of deconvolved images, $\nabla L_{i,j}(x, y)$ are the gradients of the deconvolved images, W is a local window centered at a pixel $x, y$ at which the reconstruction error is to be estimated.

The PSF scale that generates the minimum energy is considered to be the correct PSF scale and hence assigned to the corresponding pixel in PSF label map $M_i(x, y)$:

$$M_i(x, y) = \mathbf{argmin_j} \hat{E}_{i,j}(x, y). \quad (5.3)$$

The resulting scale (depth) map $M_i(x, y)$ may have some misclassified labels or gaps in the depth map. This may be due to occluded areas or regions without sufficient texture. While inpainting techniques can be used for view synthesis [119, 120], a simple approach is preferred in this work. A mode filter is applied, that is, a sliding window over the depth map to compute the mode of a local neighborhood and re-assign the center pixel to the mode label.

### 5.3.2 Blurred image deconvolution and fusion

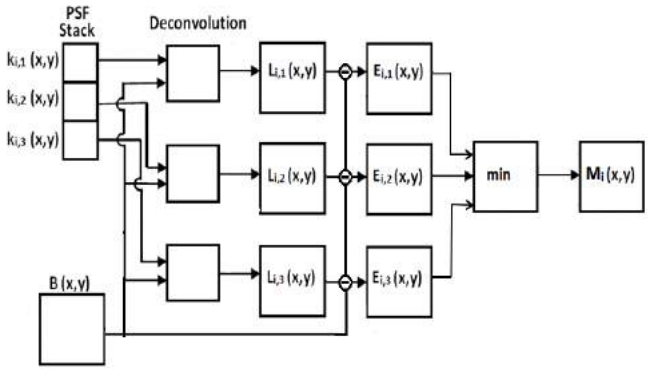



Figure 5.3: Single perspective image recovery process from the blurred image.

The recovery of entire light field involves the repetition of steps required for a single perspective image recovery shown in Figure 5.3 and 5.4, and detailed below.

First, the input image $B(x, y)$ is deconvolved for every color channel using [109] with the PSFs, known through the previous PSF scale identification step. The deconvolved image $L_{i,j}(x, y)$ has one of the depths recovered while ringing artifacts appear on regions other than the recovered depth. The depth map $M_i(x, y)$ is used to generate a set of binary masks corresponding to the PSF labels; and these masks applied on corresponding deconvolved images to construct a single perspective image $L_i(x, y)$. The process is repeated for all perspectives.

## 5.4 Experimental results

In this section, the results of light field reconstruction from a single blurred image of size 1024 x 1280 pixels are presented. In the experiments, a light fields of 11 x 11 angular resolutions is generated. The size of each recovered perspective image is again 1024 x 1280 pixels. In the experiments the target is placed at, 30cm away from the camera and moved until 2 meters from the camera with a separation of 10cm resulting in 18 different depths. The size of PSF is set to 69 x 69 pixels.

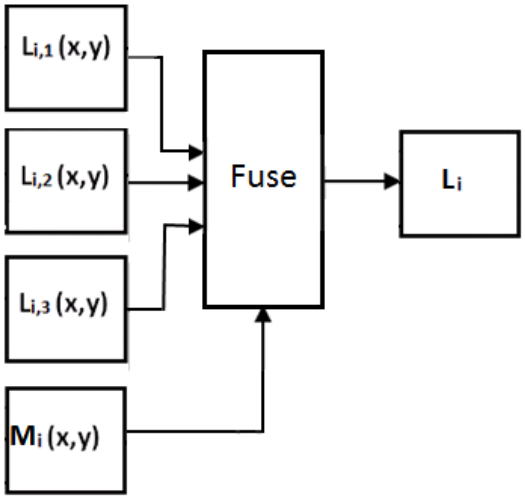



Figure 5.4: Fusion of deconvolved images to generate the final perspective image .

In Figure 5.5, a subset of the PSFs obtained in the offline calibration process are shown. In Figure 5.6, an input image $B(x, y)$, images deconvolved with the PSFs corresponding to the middle perspective and three different depths, the label map and the final perspective image are presented. From the deconvolved images, it is seen that when the PSF scale does not match the region, ringing artifacts occur. The final image is constructed from the first two deconvolved images given in the figure.

Figure 5.7, compares the central perspective image of a light field generated using the proposed algorithm with a regular image captured from same prototype camera through a pinhole on the full sensor.

In addition to the comparison of the central perspective image of the light field obtained from the proposed method with a regular image captured from same prototype camera, a comparison with central perspective image of Lytro camera's light field is also presented in Figure 5.8. The Lytro camera image is decoded by [14] which produces a regular grid of 11 x 11 perspective images with the spatial resolution of 380 x 380 pixels. The overlapping part of the Lytro's perspective image is bicubically resized to the match size of the prototype camera image.

Figure 5.9 and 5.10, presents the entire 11 x 11 light field generated using the

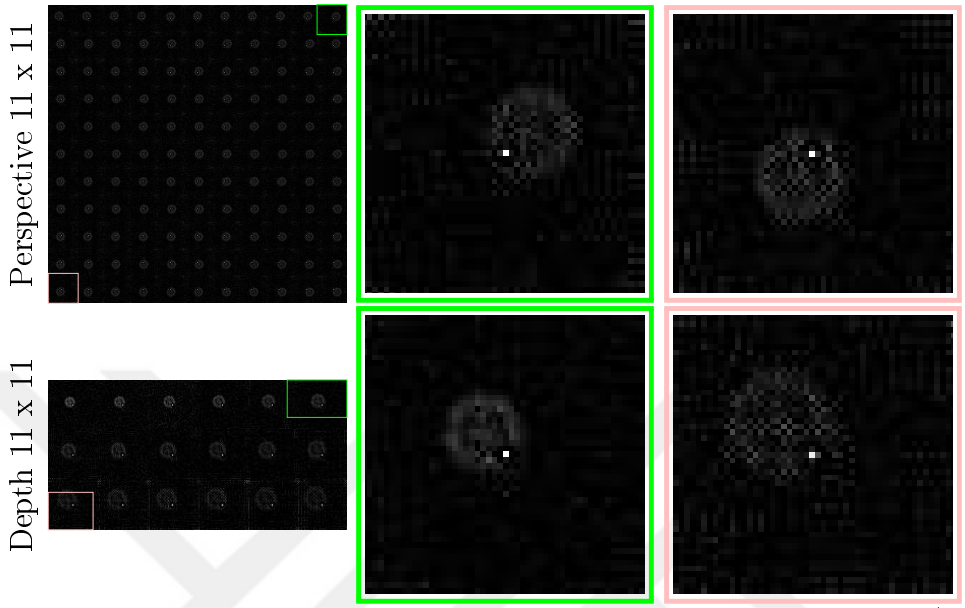



Figure 5.5: Subset of the PSFs obtained in the offline calibration process. (Row 1) PSF estimates of all perspectives at $d_j$ sampled depth of 11 x 11 light field. (Row 3) PSF stack of multiple depths of $p_i$ perspective position of 11 x 11 light field.

proposed method along with the corresponding horizontal and vertical epipolar plane images (EPIs).

To characterize the imaging performance a standard ISO 12233 test chart is used. The results in the Figure 5.11 shows the better quality of the proposed algorithm's light field over the Lytro first generation camera's light field, however when compared with regular image, resolution limitations and slight chromatic aberrations are visible.

One of the most attractive features of light field camera is its ability of post-capturing refocusing. In Figure 5.13 and 5.14 the refocusing on a high-resolution light field generated through the proposed algorithm is successfully demonstrated .

## 5.5 Discussion

In this work, an extraction of a high-resolution light field from a blurred image is successfully . The reconstruction algorithm requires a PSF label/depth map which represents different depths present in the scene. This PSF scale label/depth map like any other passive depth estimation technique relies heavily on the texture of the scene. Even though the gaps in the PSF label/depth map due to fewer texture regions are handled by the algorithm from the information of local neighborhood, The proposed algorithm at the moment does not incorporate any mechanism to deal with the fewer texture surfaces. As the depth discrimination criterion in the proposed algorithm exploits the sparsity of the natural images, its application is also limited to natural images.

Depth estimation from a single image has always been a challenging problem yet a variety of algorithms exits for this purpose and the chosen algorithm is just one of the existing methods. However with the availability of modern day computational resources and learning-based tools, in future, it is planed to overcome all these dependencies by the incorporation of a learning-based depth map estimation algorithm from a single image. In the current setup, the maximum depth range is limited to 2 meters from the camera but in future, it is planned to extend this depth range by the controlled camera motion during the exposure time to extend the effective aperture size.

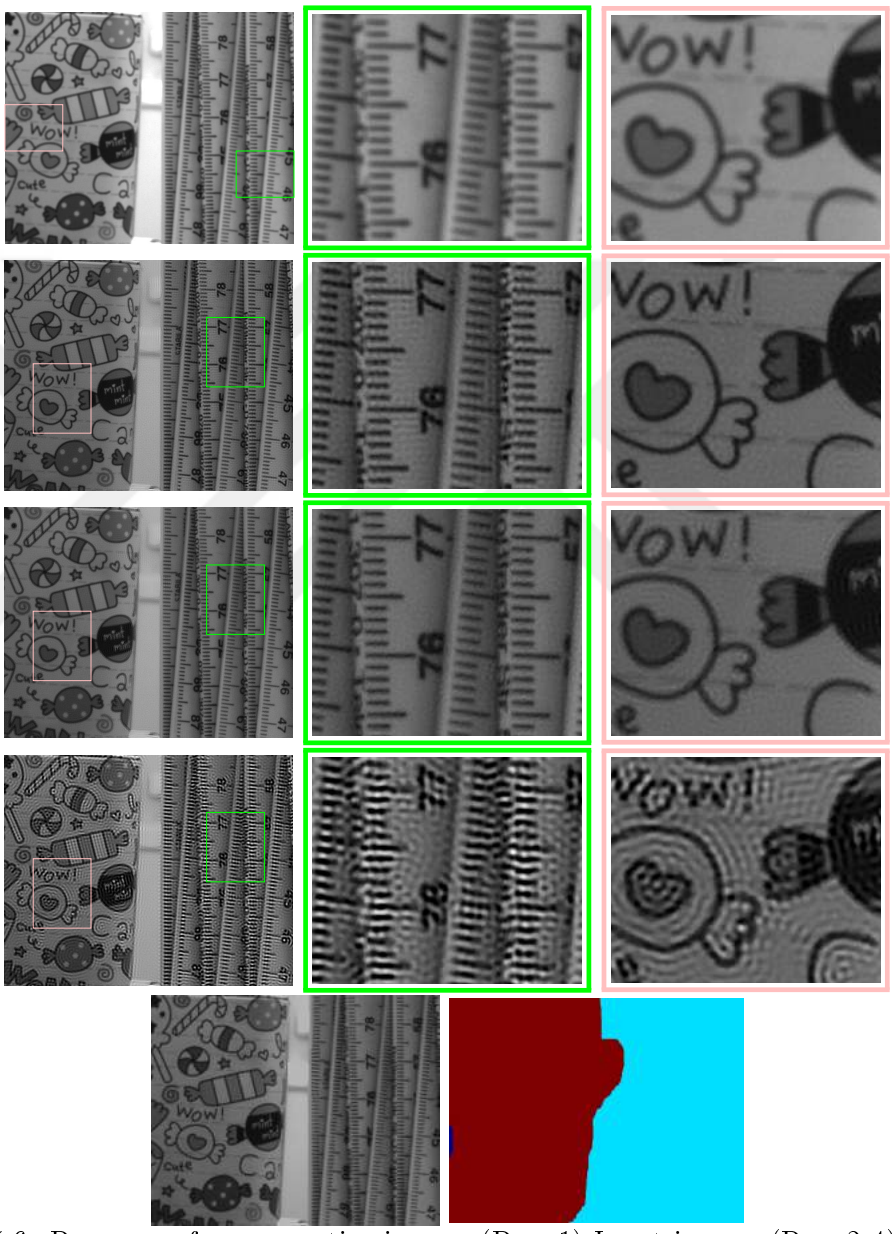


Figure 5.6: Recovery of a perspective image. (Row 1) Input image. (Row 2-4) Deconvolved images with different PSFs. The first two deconvolved images have matching depths, which can be identified from the zoomed-in regions; the last deconvolved image does not have any matching depth. (Row 5) Recovered perspective image and the label map indicating the regions taken from the first two deconvolved images. The PSFs used in the deconvolution are given in Figure 5.5.

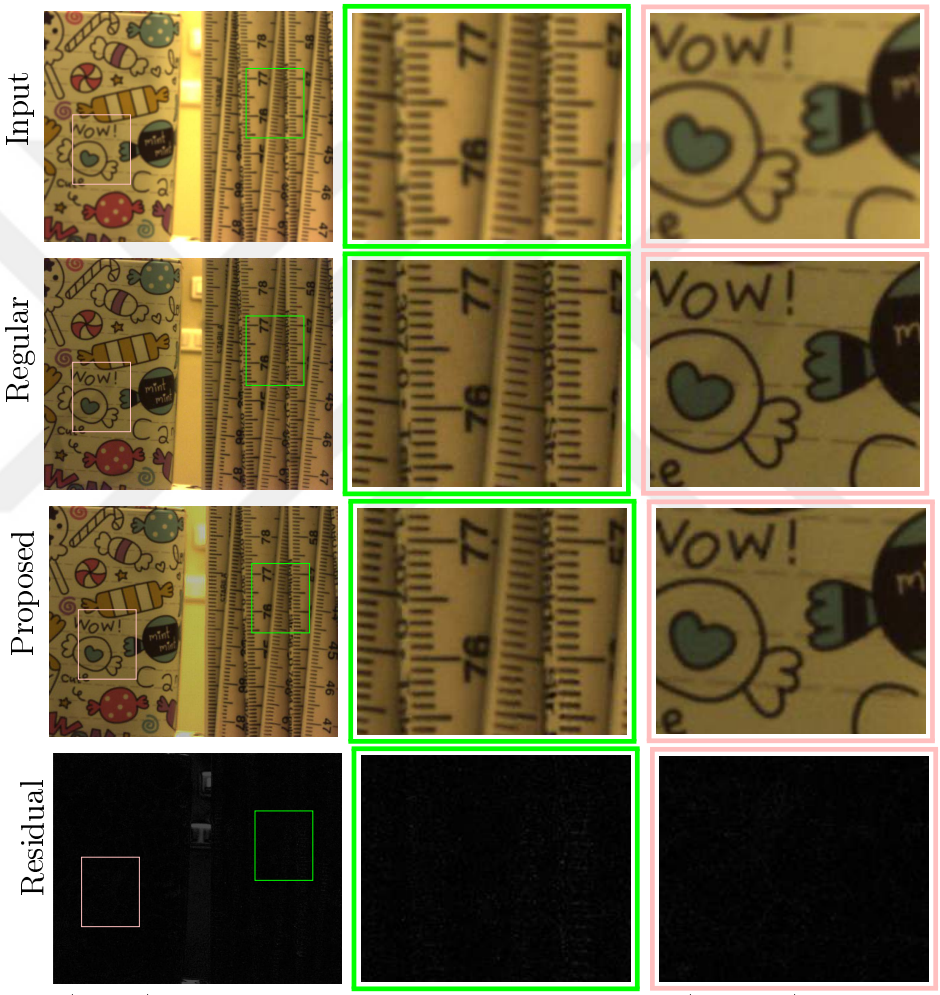



Figure 5.7: (Input) Input image to the proposed algorithm. (Regular) Image captured from prototype camera through a pinhole on the full sensor. (Proposed) Central perspective image of the light field generated using the proposed algorithm. (Residual) The residual between the regular image and central perspective image. PSNR and structure similarity between the regular image and proposed algorithm's light field central perspective image PSNR = 27.0280, SSIM = 0.94.

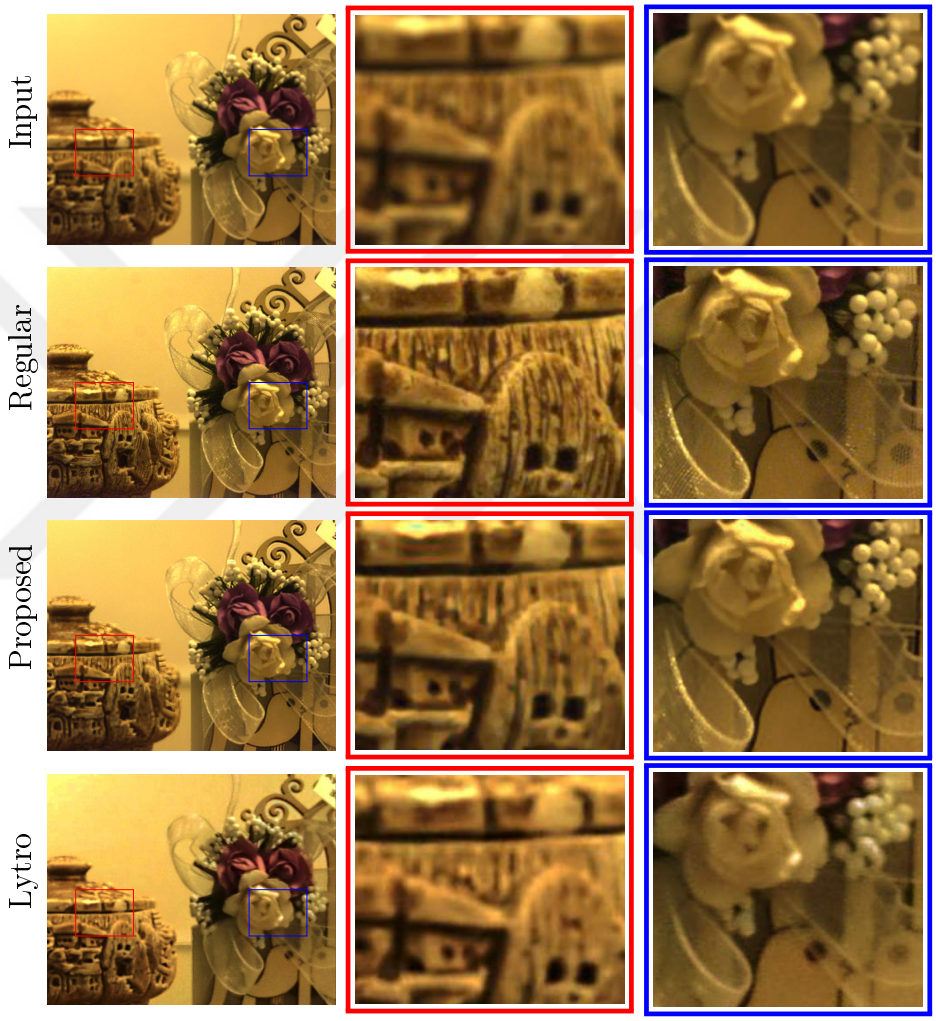



Figure 5.8: Comparison of 11 x 11 high-resolution light field recovered by the proposed method with bicubically resized Lytro's light field. (Input) Input image to the proposed algorithm. (Regular) Image captured from prototype camera through a pinhole on the full sensor. (Proposed) Central perspective image of the light field generated using the proposed algorithm. (Lytro) Central perspective image of the light field from Lytro camera. PSNR and structure similarity between the regular image and proposed algorithm's light field central perspective image PSNR = 28.4208, SSIM = 0.9571. PSNR and structure similarity between the regular image and Lytro's light field central perspective image PSNR = 25.9110, SSIM = 0.9497

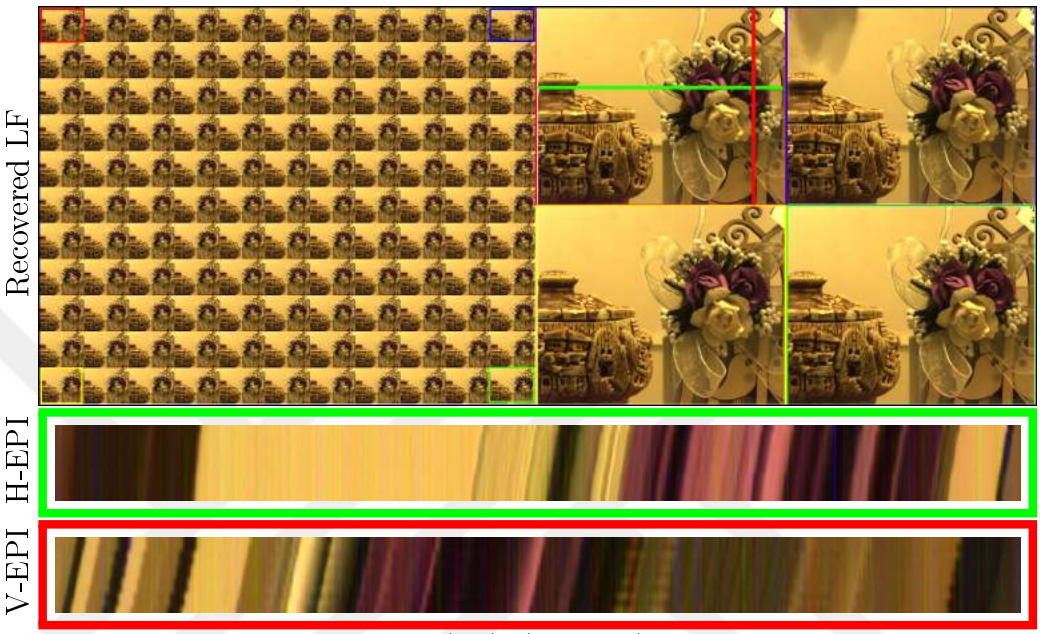



Figure 5.9: Reconstructed light field (LF). (Top row) 11 x 11 reconstructed light field and 4 zoomed-in corner perspectives. (Middle row) Horizontal EPI. (Bottom row) Vertical EPI.

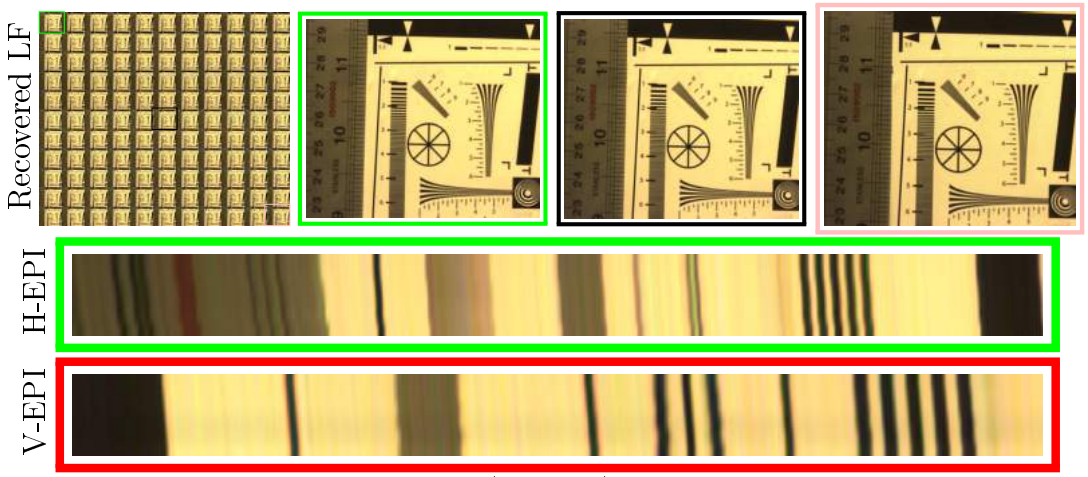



Figure 5.10: Reconstructed light field. (Top row) 11 x 11 reconstructed light field and 3 zoomed-in perspective images. (Middle row) Horizontal EPI. (Bottom row) Vertical EPI.

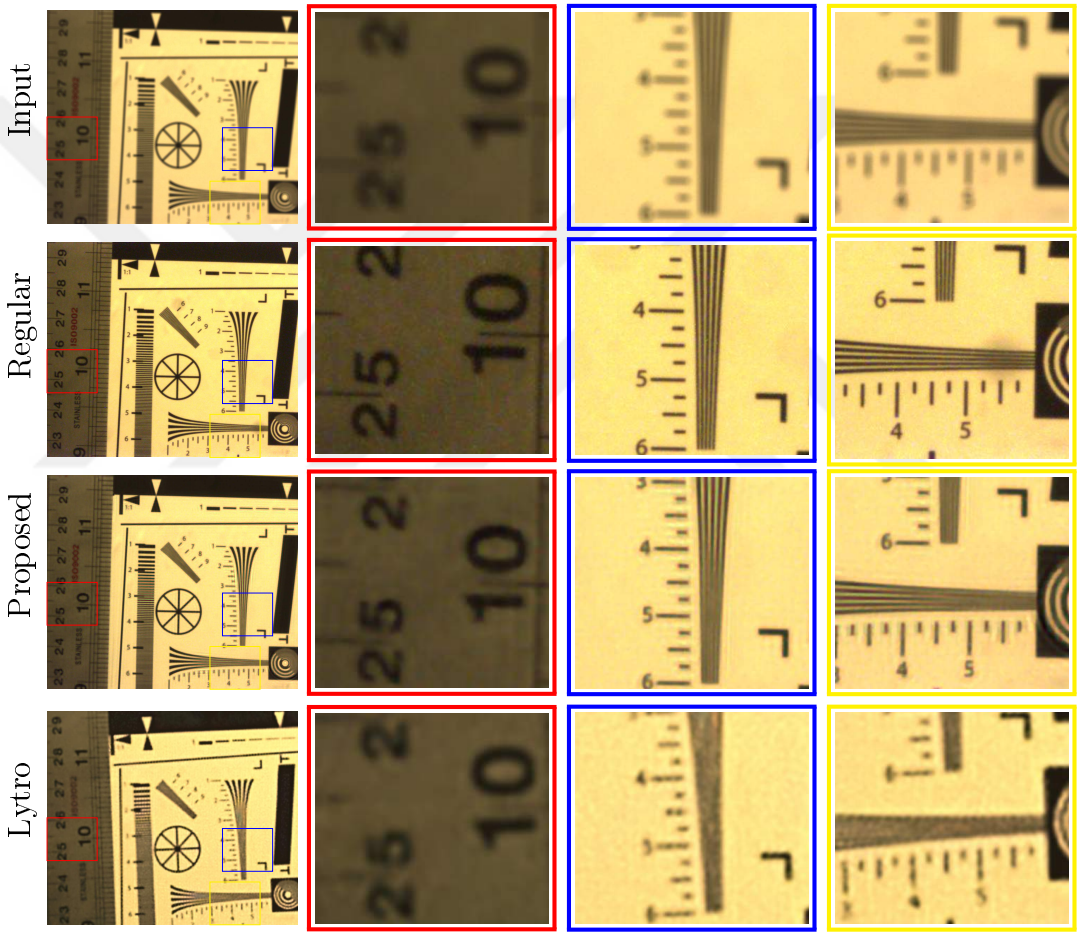



Figure 5.11: Comparison of 11 x 11 high-resolution light field recovered by the proposed method with bicubically resized Lytro's light field and regular camera image. (Input) Input image to the proposed algorithm. (Regular) Image captured from prototype camera through a pinhole on the full sensor. (Proposed) Central perspective image of the light field generated using the proposed algorithm.

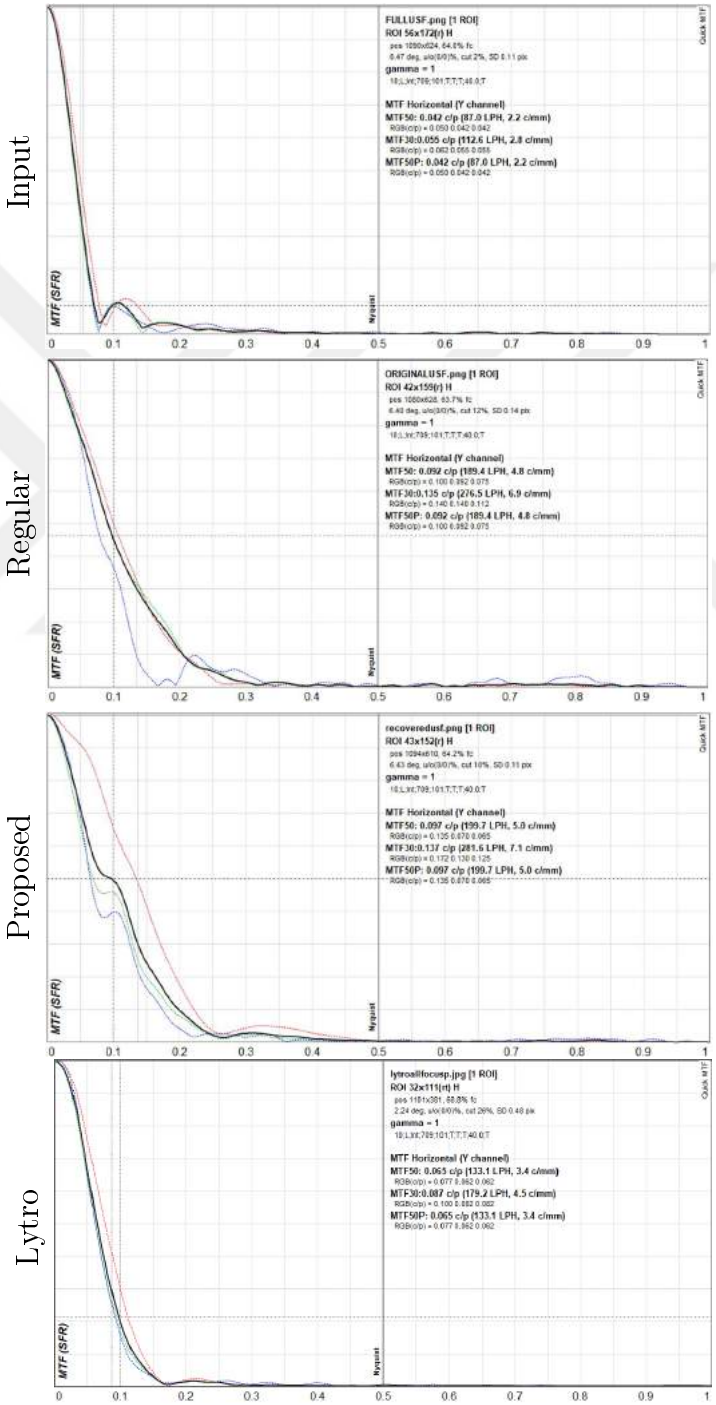



Figure 5.12: Modulation transfer function of all the images in Figure 5.11.

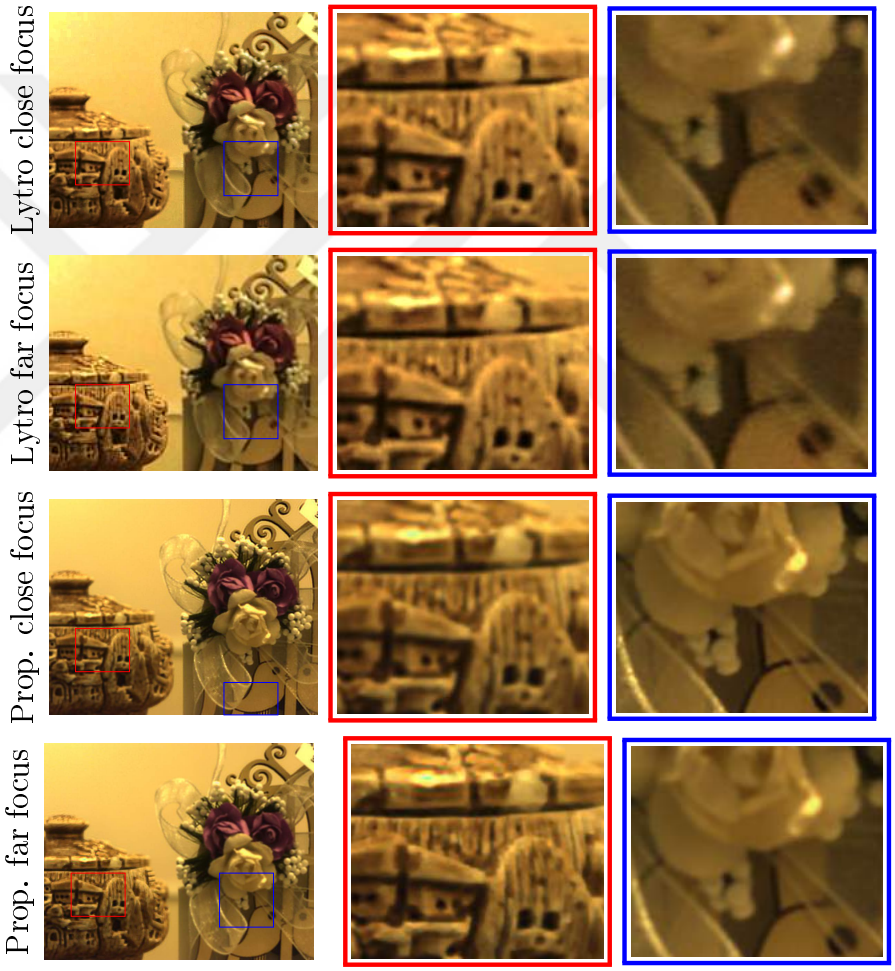



Figure 5.13: Comparison of post-capture refocusing of high-resolution 11 x 11 light field. (Prop.) Recovered by the proposed method. (Lytro) A bicubically resized Lytro's light field.

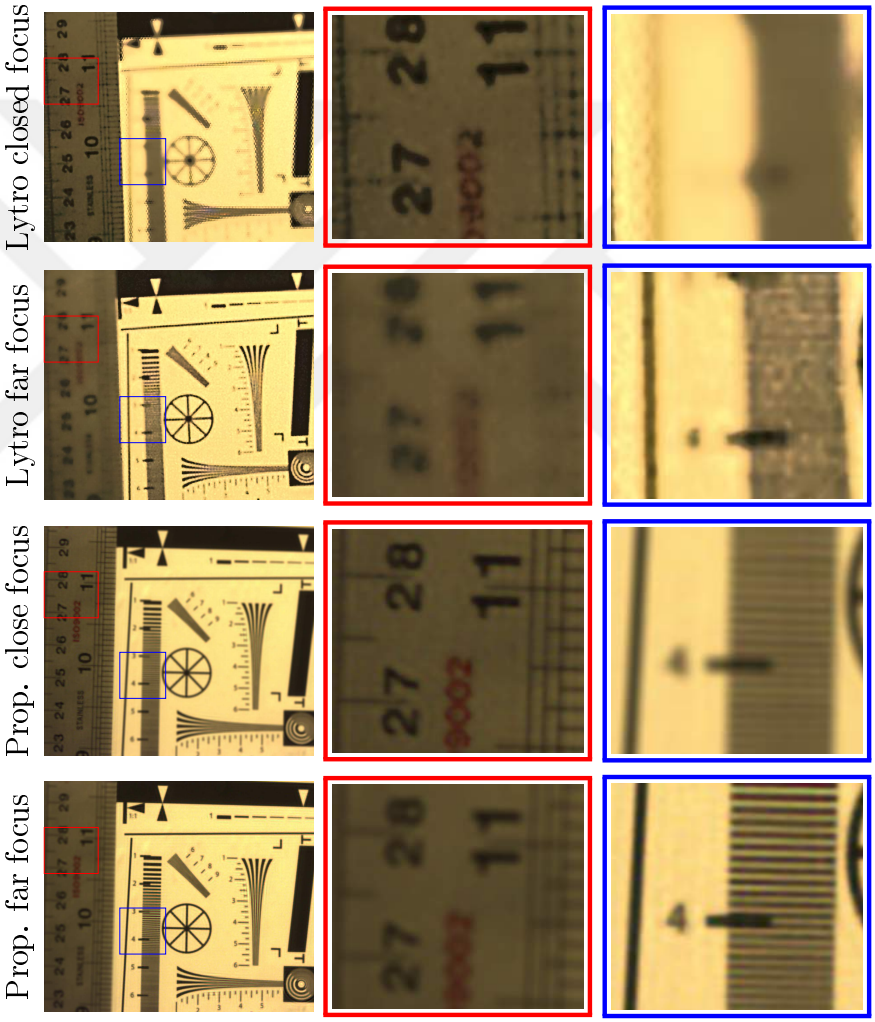



Figure 5.14: Comparison of Post-capture refocusing of high-resolution 11 x 11 light field. (Prop.) Recovered by the proposed method. (Lytro) A bicubically resized Lytro's light field.

# Chapter 6

# Conclusion

In this thesis, few limitations of the conventional imaging system which results in the loss of quality of the image are addressed. Through computational imaging techniques, we overcome these limitations and explored new possibilities for digital images. The major limitations addressed in this work are limited dynamic range, limited depth of field, space-variant blurring and limited spatial resolution in the context of light field imaging.

First, a blind image deblurring method for space-variant blur is developed. The method can handle space-variant blur of any type without any parametric assumption of the movement of the camera or the objects. The method splits the image into overlapping patches and estimates the blur kernel for each patch. The estimated kernels are then clustered to find the main blur kernels and the corresponding regions. Some regions may remain unassigned to a cluster. For each region, the blur kernel is refined. The input image is then deblurred with each refined blur kernel. An image fusion process merges these deblurred images into a single one, picking the optimum pixel using an energy function. While the provided results are obtained with a specific set of techniques for each step, the framework can be used with other techniques and parameter choices, making it possible to improve in the future. The method is compared with various other space-variant deblurring methods; and it produces comparable or better results

on a variety of blur scenarios, including camera shake, out-of-focus blur, and object motion.

In the proposed method for space-variant image deblurring, we try to model large abrupt blur variations over the entire image as small patches of smoothly varying or uniform blur within a patch and large changes between the patches. Also, we assume that patch size should be large enough to have a rich texture for accurate PSF estimation. It is because of these assumptions, the size of the patch becomes very critical. In addition, the use of overlapping window requires determination of an adequate shift amount, which can neither be too small, due to computational complexity, nor too large to produce fewer samples. Finally, the bandwidth of the clustering algorithm is also very critical in generating accurate clusters and that too along with the patch size and stride amount is set empirically in the current implementation. In the future, we would like to investigate it further and plan to come up with an adaptive method for the selection of these patches based on certain criteria, for example, the availability of texture within each patch.

In the current implementation, neither object segmentation nor boundary conditions are kept in consideration during image segregation into regions with distinct blur. To make our algorithm applicable to more general types of space-variant blur, pre-segmentation for separating regions of distinct blurs in an image can be adapted in future.

Secondly, an algorithm is proposed for extending the depth of field and dynamic range from differently focused and exposed images captured through a camera array. The core process is focus stacking regardless of the exposure value; it requires photometric and spatial registration and includes pixel-by-pixel weight calculation (involving sharpness, saturation, and registration errors) for fusion. For HDR imaging, it is proposed to perform focus stacking for each exposure category and fuse the focus-stacked images using a standard HDR imaging algorithm. The experiments are done with real data and satisfactory results are obtained. There is some future work that may improve results. The choice of the optical flow algorithm is critical and should be investigated further as spatial registration

errors may lead to artifacts. The appropriate choice of optical flow algorithm can enable the possibility for the registration of differently focused images and hence a more effective utilization of the camera array can be achieved.

Finally, two different camera setups are designed to overcome the spatial resolution problem with hand held light field cameras. A hybrid imaging system that includes a light field camera and a regular camera is proposed. The system, while keeping the capabilities of light field imaging, improves both spatial resolution and depth estimation range/accuracy due to increased baseline. Because the fixed stereo system allows pre-calibration and by utilizing the fact that light field sub-aperture images are captured on a regular grid, the registration of low-resolution light field sub-aperture images and high-resolution regular camera images is simplified. With proper image registration, even a simple image fusion, such as alpha blending, produces good results.

Another method is proposed to recover light field from a single capture with a regular camera. The method is based on deconvolving a wide aperture image with a set of PSFs calculated in an offline calibration process. It has good light efficiency and can achieve spatial resolution as much as the camera sensor has since it does not have any spatio-angular trade-off, unlike the MLA based light field cameras. The method can be used to convert any regular camera with a controllable aperture into a light field camera.

In the current implementation, the segmentation algorithm used requires user interaction, to define the number of depths in the image. User interaction is needed only once per light field extraction. However, Image segmentation is a well-investigated topic in computer vision and a variety of both conventional and learning base segmentation methods exists in the literature. Although there are many learning based methods which produce good results without any user interaction, but most of these algorithms are pre-trained only for the outdoor environment. Methods designed for indoors will still require fine-tuning with sufficiently large datasets to produce acceptable results. In the future, we would like to investigate the automatic segmentation algorithms to perform segmentation without human interaction.

Overall in this work, four different algorithms based on computational imaging techniques are developed which address a number of limitation in the field of computer vision. The work enables several new application and improves the existing imaging applications, ranging from improved image classification to better depth estimation and post-capture refocusing to high dynamic range imaging. Every algorithm developed addresses some of the challenging research topics in computational photography and has still sufficient room for improvement. We hope to address the above-mentioned limitations in the future to make these algorithms more robust and user-friendly and eventually transform these into software packages from commercial usage.

# Bibliography


[1] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf, "Fast removal of non-uniform camera shake," in *IEEE Int. Conf. on Computer Vision*, vol. 463–470, November 2011.

[2] T. H. Kim and K. M. Lee, "Segmentation-free dynamic scene deblurring," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2766-2773, June 2014.

[3] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "Image reconstruction from a single exposure using deep cnns," *ACM Transactions on Graphics*, vol. 36, 2017.

[4] M. Levoy and P. Hanrahan, "Light field rendering," in *ACM Int. Conf. on Computer Graphics and Interactive Techniques*, pp. 31–42, 1996.

[5] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics.*, vol. 24, pp. 765–776, 2005.

[6] R. Ng, "Fourier slice photography," *ACM Transactions on Graphics*, vol. 24, pp. 735–744, 2005.

[7] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving resolution and depth-of-field of light field cameras using a hybrid imaging system," in *IEEE Int. Conf. on Computational Photography*, pp. 1–10, 2014.

[8] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Transactions on Graphics*, vol. 32, pp. 46:1–46:12, 2013.

[9] Z. Hu, L. Xu, and M. H. Yang, "Joint depth estimation and camera shake removal from single blurry image," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2893-2900, June 2014.

[10] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski, "Image deblurring using inertial measurement sensors," *ACM Transactions on Graphics*, vol. 29, 2010.

[11] S. Nah, H. T. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 3883-3891, July 2017.

[12] W. S. Lai, J. B. Huang, Z. Hu, N. Ahuja, and M. H. Yang, "A comparative study for single image blind deblurring," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1701–1709, June 2016.

[13] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 472–479, 2000.

[14] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1027–1034, 2013.

[15] D. Cho, M. Lee, S. Kim, and Y. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *IEEE Int. Conf. on Computer Vision*, pp. 3280–3287, 2013.

[16] F. C. Calderon, C. A. Parra, and C. L. Niño, "Depth map estimation in light fields using an stereo-like taxonomy," in *IEEE Conf. Signal Processing and Artificial Vision*, pp. 1–5, 2014.

[17] A. Gershun, "The light field," *Journal of Mathematics and Physics*, vol. 18, no. 1, pp. 51–151, 1939.

[18] E. H. Adelson and J. R. Bergen, *The plenoptic function and the elements of early vision*. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1991.

[19] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *ACM Conf. on Computer Graphics and Interactive Techniques*, pp. 43–54, 1996.

[20] B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics*, vol. 24, pp. 765–776, 2005.

[21] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, "A real-time distributed light field camera," in *Eurographics Workshop on Rendering*, pp. 77–86, 2002.

[22] A. Lumsdaine and T. Georgiev, "The focused plenoptic camera," in *IEEE Int. Conf. on Computational Photography*, pp. 1–8, 2009.

[23] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Transactions on Graphics*, vol. 26, 2007.

[24] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala, "Spatio-angular resolution tradeoffs in integral photography," in *Eurographics Conf. on Rendering Techniques*, pp. 263–272, 2006.

[25] J. Unger, A. Wenger, T. Hawkins, A. Gardner, and P. Debevec, "Capturing and rendering with incident light fields," in *Eurographics Workshop on Rendering*, pp. 141–149, 2003.

[26] "Lytro, inc." `https://support.lytro.com/hc/en-us/`. Accessed: 2018-05-24.

[27] "Raytrix, gmbh." `https://www.raytrix.de/`. Accessed: 2018-05-24.

[28] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu, "Line assisted light field triangulation and stereo matching," in *IEEE Int. Conf. on Computer Vision*, pp. 2792–2799, 2013.

[29] H. G. Jeon, J. Park, G. Choe, and J. Park, "Accurate depth map estimation from a lenslet light field camera," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1547–1555, 2015.

[30] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *IEEE Int. Conf. on Computer Vision*, pp. 673–680, 2013.

[31] C. K. Liang, T. H. Lin, B. Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," *ACM Transactions on Graphics*, vol. 27, pp. 55:1–55:10, 2008.

[32] A. Ashok and M. A. Neifeld, "Compressive light field imaging," in *SPIE Defense, Security, and Sensing*, vol. 7690, 2010.

[33] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive light field sensing," *IEEE Trans. on Image Processing*, vol. 21, pp. 4746 – 4757, 2012.

[34] Y. P. Wang, L. C. Wang, D. H. Kong, and B. C. Yin, "High-resolution light field capture with coded aperture," *IEEE Trans. on Image Processing*, vol. 24, pp. 5609 – 5618, 2015.

[35] R. Fergus, B. Singh, A. Hertzmann, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Transactions on Graphics*, vol. 25, pp. 787–794, July 2006.

[36] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Transactions on Graphics*, vol. 27, pp. 1–10, August 2008.

[37] N. Joshi, R. Szeliski, and D. J. Kriegman, "PSF estimation using sharp edge prediction," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1-8, June 2008.

[38] S. Cho and S. Lee, "Fast motion deblurring," *ACM Transactions on Graphics*, vol. 28, pp. 145:1–145:8, December 2009.

[39] L. Xu and J. Y. Jia, "Two-phase kernel estimation for robust motion deblurring," *European Conf. on Computer vision*, vol. 6311, pp. 157–170, September 2010.

[40] J. Cai, H. Ji, C. Liu, and Z. Shen, "Blind motion deblurring from a single image using sparse approximation," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 104-111, June 2009.

[41] J. Chen, L. Yuan, C. K. Tang, and L. Quan, "Robust dual motion deblurring," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1-8, June 2008.

[42] F. Li, J. Y. Yu, and J. X. Chai, "A hybrid camera for motion deblurring and depth map super-resolution," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1-8, June 2008.

[43] L. Yuan, J. Sun, L. Quan, and H. Y. Shum, "Image deblurring with blurred/noisy image pairs," *ACM Transactions on Graphics*, vol. 26, July 2007.

[44] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 98, pp. 168–186, June 2010.

[45] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," *European Conf. on Computer Vision*, vol. 6311, pp. 171–184, September 2010.

[46] O. Šindelář and F. Šroubek, "Image deblurring in smartphone devices using built-in inertial measurement sensors," *Journal of Electronic Imaging*, vol. 22, February 2013.

[47] Q. Shan, W. Xiong, and J. Y. Jia, "Rotational motion deblurring of a rigid object from a single image," in *IEEE Int. Conf. on Computer Vision*, vol. 1-8, October 2007.

[48] J. Jia, "Single image motion deblurring using transparency," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, June 2007.

[49] Y. W. Tai, H. X. Tang, M. S. Brown, and S. Lin, "Detail recovery for single image defocus blur," *Transactions on Computer Vision and Applications*, vol. 1, pp. 95–104, March 2009.

[50] Y. W. Tai and M. S. Brown, "Single image defocus map estimation using local contrast prior," in *IEEE Int. Conf. on Image Processing*, vol. 1797-1800, November 2009.

[51] C. T. Shen, W. L. Hwang, and S. C. Pei, "Spatially-varying out-of-focus image deblurring with L1-2 optimization and a guided blur map," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1069-1072, March 2012.

[52] H. Cheong, E. Chae, E. Lee, G. Jo, and J. Paik, "Fast image restoration for spatially-varying defocus blur of imaging sensor," *Sensors*, vol. 15, pp. 880–898, December 2015.

[53] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics*, vol. 26, pp. 1–9, July 2007.

[54] R. Liu, Z. Li, and J. Jia, "Image partial blur detection and classification," vol. 1-8, June 2008.

[55] A. Chakrabarti, T. Zickler, and W. T. Freeman, "Analyzing spatially-varying blur," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2512-2519, June 2010.

[56] A. Levin, "Blind motion deblurring using image statistics," *Advances in Neural Information Processing Systems*, vol. 841-848, December 2006.

[57] Z. Shen, T. Xu, J. Pan, and J. Guo, "Non-uniform motion deblurring with kernel grid regularization signal process," *Signal Processing: Image Communication*, vol. 62, pp. 1–15, 2018.

[58] S. Harmeling, H. Michael, and B. Schoelkopf, "Space-variant single image blind deconvolution for removing camera shake," *Advances in Neural Information Processing Systems*, vol. 1, pp. 829–837, December 2010.

[59] H. Ji and K. Wang, "A two-stage approach to blind spatially-varying motion deblurring," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 73–80, June 2012.

[60] J. Sun, W. Cao, Z. Xu, and Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 769-777, March 2015.

[61] Q. Qian and B. K. Gunturk, "Space-varying blur kernel estimation and image deblurring," *IS&T/SPIE Electronic Imaging*, vol. 9023, February 2014.

[62] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790–799, August 1995.

[63] J. F. Wang, Y. Zhang, and W. T. Yin, "An efficient TVL1 algorithm for deblurring multi-channel images corrupted by impulsive noise," *Journal on Scientific Computing*, vol. 31, pp. 2842–2865, July 2009.

[64] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High-speed videography using a dense camera array," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. II–II, 2004.

[65] V. Vaish, M. Levoy, R. Szeliski, C. L. Zitnick, and S. B. Kang, "Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 2331–2338, 2006.

[66] T. Li and D. P. K. Lun, "Super-resolution imaging with occlusion removal using a camera array," in *IEEE Int. Symposium on Circuits and Systems*, pp. 2487–2490, 2016.

[67] Y. Tai, H. Du, M. S. Brown, and S. Lin, "Correction of spatially varying image and video motion blur using a hybrid camera," *IEEE Transactions*

*on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1012–1028, 2010.

[68] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 472–479, 2000.

[69] D. D. Wen, "High dynamic range charge-coupled device," in *US Patent 4873561*, 1989.

[70] R. A. Street, "High dynamic range segmented pixel sensor array," in *US Patent 5789737*, 1998.

[71] J. Tumblin, A. Agrawal, and R. Raskar, "Why i want a gradient camera," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 103–110, 2005.

[72] S. Mann and R. W. Picard, "On being undigital with digital cameras: Extending dynamic range by combining differently exposed pictures," in *IEEE Int. Workshop on Imaging Systems and Techniques*, pp. 442–448, 1995.

[73] A. ab, M. Z. Alam, and B. K. Gunturk, "High dynamic range imaging using a plenoptic camera," in *IEEE Int. Conf. Signal Processing and Communications Applications Conference*, pp. 1–4, 2017.

[74] T. Georgiev, "New results on the plenoptic 2.0 camera," in *Asilomar Conf. on Signals, Systems, and Computers*, pp. 1243–1247, 2009.

[75] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Intl. Conf. on Computer Graphics and Interactive Techniques*, pp. 369–378, 1997.

[76] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 374–380 Vol. 1, 1999.

[77] J. Tian and L. Chen, "Multi-focus image fusion using wavelet-domain statistics," in *IEEE Int. Conference on Image Processing*, pp. 1205–1208, 2010.

[78] M. Antunes, M. Trachtenberg, G. Thomas, and T. Shoa, "All-in-focus imaging using a series of images on different focal planes," in *Image Analysis and Recognition*, pp. 174–181, 2005.

[79] Z. Z. and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proceedings of the IEEE*, vol. 87, pp. 1315–1326, 1999.

[80] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 3, pp. 266–274, 1995.

[81] K. Kodama, H. Mo, and A. Kubota, "Simple and fast all-in-focus image reconstruction based on three-dimensional/two-dimensional transform and filtering," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–769–I–772, 2007.

[82] F. Aguet, D. V. D. Ville, and M. Unser, "Model-based 2.5-d deconvolution for extended depth of field in brightfield microscopy," *IEEE Transactions on Image Processing*, vol. 17, pp. 1144–1153, 2008.

[83] B. Forster, D. Ville D. V. D, J. Berent, D. Sage, and M. Unser, "Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images," vol. 65, p. 33—42, 2004.

[84] C. Liu, "Beyond pixels: Exploring newrepresentations and applications formotion analysis.," in *Doctoral Thesis. Massachusetts Institute of Technology*, 2009.

[85] P. Sen., N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "' Patch-Based HDR Reconstruction of Dynamic Scenes," *ACM Transactions on Graphics*, vol. 31, pp. 203:1–203:11, 2012.

[86] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *IEEE Int. Conf. on Computational Photography*, pp. 1–9, 2009.

[87] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch

prior," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 22–28, 2012.

[88] S. Wanner and B. Goldluecke, "Spatial and angular variational super-resolution of 4d light fields," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 901–908, 2012.

[89] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *IEEE Int. Conf. Computer Vision Workshop*, pp. 57–65, 2015.

[90] N. K. Kalantari, T. C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," in *IEEE Int. Conf. Computer Vision Workshop*, pp. 997–1013, 2016.

[91] S. Wanner and B. Goldluecke, "Fourier slice super-resolution in plenoptic cameras," in *IEEE Int. Conf. on Computational Photography*, pp. 1–11, 2012.

[92] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous fourier domain," in *ACM Transactions on Graphics*, pp. 22–28, IEEE, 2012.

[93] S. A. Shroff and K. Berkner, "Image formation analysis and high resolution image reconstruction for plenoptic imaging systems," *Aplied Optics*, vol. 52, pp. 22–31, 2013.

[94] M. Broxton, L. Grosenick, S. Yang, N. Cohen, A. Andalman, K. Deisseroth, and M. Levoy, "Wave optics theory and 3d deconvolution for the light field microscope," *Optics Express*, vol. 21, pp. 25418–25439, 2013.

[95] J. M. Trujillo-Sevilla, L. F. Rodriguez-Ramos, I. Montilla, and J. M. Rodriguez-Ramos, "High resolution imaging and wavefront aberration correction in plenoptic systems," *Optics Letters*, vol. 39, pp. 5030–5033, 2014.

[96] A. Junker, T. Stenau, and K. H. Brenner, "Scalar wave-optical reconstruction of plenoptic camera images," *Applied Optics*, vol. 53, pp. 5784–5790, 2014.

[97] T. Georgiev, "New results on the plenoptic 2.0 camera," in *Asilomar*, pp. 1243–1247, 2009.

[98] J. Wu, H. Wang, X. Wang, and Y. Zhang, "A novel light field super-resolution framework based on hybrid imaging system," in *Visual Communications and Image Processing*, pp. 1–4, 2015.

[99] X. Wang, L. Li, and G. Houi, "High-resolution light field reconstruction using a hybrid imaging system," *Applied Optics*, vol. 55, no. 10, pp. 2580–2593, 2016.

[100] D. Gallup, J. M. Frahm, P. Mordohai, and M. Pollefeys, "Variable baseline/resolution stereo," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[101] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4d light fields," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 41–908, 2012.

[102] M. D. Grossberg and S. K. Nayar, "Determining the camera response from images: what is knowable?," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1455–1467, 2003.

[103] C. Liu, "Beyond pixels: exploring new representations and applications for motion analysis," MIT, 2009.

[104] H. B. Mitchell, *Image fusion: theories, techniques and applications*. Springer, 2010.

[105] P. M. Zeeuw, "Wavelet and image fusion," in *CWI*, 1998.

[106] V. Aslantas, "A depth estimation algorithm with a single image," *Optics Express*, vol. 15, pp. 5024–5029, 2007.

[107] J. Lin, X. Ji, W. Xu, and Q. Dai, "Absolute depth estimation from a single defocused image," *IEEE Trans. on Image Processing*, vol. 22, pp. 4545 – 4550, 2013.

[108] X. Zhu, S. Cohen, S. Schiller, and P. Milanfar, "Estimating spatially varying defocus blur from a single image," *IEEE Trans. on Image Processing*, vol. 22, pp. 4879–4891, 2013.

[109] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics*, vol. 26, 2007.

[110] A. Sellent and P. Favaro, "Optimized aperture shapes for depth estimation," *Pattern Recognition Letters*, vol. 40, pp. 96–103, 2014.

[111] M. Masoudifar and H. R. Pourreza, "Image and depth from a single defocused image using coded aperture photography," *CoRR*, 2016.

[112] A. Veeraghavan, "Dappled photography : Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Transactions on Graphics*, vol. 26, pp. 1–12, 2007.

[113] C. Zhou and S. Nayar, "What are good apertures for defocus deblurring?," in *IEEE Int. Conf. on Computational Photography*, pp. 1–8, 2009.

[114] M. Belen, P. Lara, C. Adrian, and G. Diego, "Perceptually optimized coded apertures for defocus deblurring," *Comput. Graph. Forum*, vol. 31, pp. 1867–1879, 2012.

[115] C. Zhou, S. Lin, and S. Nayar, "Coded aperture pairs for depth from defocus," in *IEEE Int. Conf. on Computer Vision*, vol. 14, pp. 99–106, 2009.

[116] Y. Takeda, S. Hiura, and K. Sato, "Fusing depth from defocus and stereo with coded apertures," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 209–216, 2013.

[117] W. M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *IEEE Int. Conf. on Computer Vision*, pp. 673 – 680, 2013.

[118] J. Lin, X. Lin, X. Ji, and Q. Dai, "Separable coded aperture for depth from a single image," *IEEE Signal Processing Letters*, vol. 21, pp. 1471–1475, 2014.

[119] I. Daribo and B. Pesquet-Popescu, "Depth-aided image inpainting for novel view synthesis," in *IEEE Int. Workshop on Multimedia Signal Processing*, pp. 167–170, 2010.

[120] K. Oh, S. Yea, and Y. Ho, "Hole filling method using depth based inpainting for view synthesis in free viewpoint television and 3D video," in *IEEE Picture Coding Symposium*, pp. 1–4, 2009.

# IMAGE ENHANCEMENT THROUGH NEW TECHNIQUES IN COMPUTATIONAL PHOTOGRAPHY

ORIGINALITY REPORT

| 12% | 7% | 10% | 1% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ www.cv-foundation.org

Internet Source

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | On | | |