Universidad de la República
Facultad de Ciencias

Multi-focus Computational Optical Imaging in Fourier Domain

MSc. Julia Alonso

Tesis presentada para completar los requerimientos del grado de Doctor en Física

Orientador: Dr. José Ferrari

defendida el 1º de noviembre de 2016 ante el tribunal integrado por:

Dr. Daniel Ariosa (Presidente)
Dr. Raúl Donangelo
Dr. Jorge Flores
Dr. Ismael Núñez
Dr. José Ferrari

Acknowledgements

First I want to thank my advisor, Dr. José Ferrari, for his expertise, and guidance throughout the course of my PhD work. I appreciate all his contributions and time.

I am also thankful to all the members of the Applied Optics Group with whom I have worked and shared many ideas as well as to my colleagues at the Institute of Physics (IFFI) at the Engineering Faculty (UdelaR).

I would also like to express my gratitude to Federico and Leonardo, my colleagues in the Electromagnetism course, who substituted me in those lectures when I was traveling for conferences or at the last stages of the preparation of this manuscript. I really appreciate your support.

Many thanks to Amelia for the time and attention she provides every day which makes her much more than a secretary to many of us. I also thank Liber who is always there when something goes wrong with my computer. Thanks also go to Antonio and Joaquín for their work in the fine mechanical workshop, whose help with all kind of devices that might be needed in the lab is fundamental. I would also like to thank Jimena and Ana from *Programa de Desarrollo de las Ciencias Básicas* (PEDECIBA) Física for their help with bureaucracy.

I am truly grateful to my committee members for their time, interest, and helpful comments on my work: Dr. Daniel Ariosa, Dr. Raúl Donangelo, Dr. Ismael Núñez and Dr. Jorge Flores.

I very much appreciate my family and close friends and I would like to thank you all for being there along the years. A special thank goes to my mother, who I think would have been very happy to see me finish the PhD she saw me started. I miss you.

And most of all, I would like to thank Ariel. Words cannot express how grateful I am to him for supporting me and encouraging me in all my pursuits. I love you. And last but not least, I would also like to thank my dog Gala for all the walks in the park.

Abstract

Computational Optical Imaging is a fast-evolving interdisciplinary research field in which optical imaging acquisition and image processing algorithms converge to play an integral role in the reconstruction of images with novel features of interest, which are absent in the images acquired directly through the optical system. From microscopy to general photography, in many situations, the optical imaging system limits the depth-of-field of a scene, which makes it impossible to acquire the three-dimensional scene entirely in focus from a single-shot image. A possible solution to overcome this limitation, is to acquire a multi-focus image stack and then computationally reconstruct the all-in-focus image (extended depth-of-field).

In this context, we implement an optical set up for image acquisition with shallow depth-of-field, consisting in a conventional monocular camera combined with an electrically tunable lens, which does not require registration between the acquired images. Then, we propose a method based in optical image formation, that enables not only post-capture reconstruction of extended depth-of-field (all-in-focus image), but also the reconstruction of novel viewpoints of the scene, including stereoscopic pairs of images and scene refocusing with synthetically reshaped apertures. The approach is based on the reorganization of the acquired visual information considering a depth-variant point-spread-function (PSF), which, unlike depth-invariant approaches, can handle severely defocused (shallow depth-of-field) multi-focus image stacks. The method is performed in Fourier domain, without segmentation of the focused regions or estimation of the depth map, which usually introduce inaccuracies into the reconstruction.

Keywords: *Three-dimensional image processing, Fourier optics, Computational photography and imaging techniques, Image reconstruction techniques*

Resumen

La Imaginería Óptica Computacional (Computational Optical Imaging) constituye un campo de investigación interdisciplinario de rápida evolución, donde los sistemas ópticos para la adquisición de imágenes y los algoritmos de procesamiento de imágenes convergen para jugar un rol integral en la reconstrucción de imágenes con nuevas características de interés, ausentes en las imágenes que son adquiridas directamente por medio del sistema óptico. Desde la microscopía a la fotografía en general, en muchas situaciones, el sistema óptico limita la profundidad de campo de una escena y esto hace imposible adquirir una imagen de la escena tridimensional completamente en foco. Una posible solución para superar esta limitación, consiste en adquirir una secuencia de imágenes multifoco y luego reconstruir computacionalmente una imagen donde toda la escena resulte en foco (profundidad de campo extendida).

En este contexto, implementamos un dispositivo óptico para la adquisición de imágenes con poca profundidad de campo, constituido por una cámara monocular combinada con una lente de foco ajustable eléctricamente, que no requiere registro entre las imágenes adquiridas. Luego, proponemos un método basado en la formación óptica de imágenes, que permite luego de la adquisición obtener, no sólo la reconstrucción de una imagen con la profundidad de campo extendida, sino también la reconstrucción de nuevos puntos de vista de la escena, incluyendo pares estereoscópicos de imágenes y reenfoque de la escena a través de aperturas reconfiguradas sintéticamente. El acercamiento propuesto se basa en la reorganización de la información visual adquirida, considerando una respuesta al impulso (PSF) variable en profundidad, la cual, a diferencia de los acercamientos invariantes en profundidad, puede tratar con secuencias de imágenes multifoco con gran desenfoque (poca profundidad de campo). El método es implementado en el dominio de Fourier, sin necesidad de segmentación de las regiones en foco o la estimación del mapa de profundidad, lo cual generalmente introduce imprecisiones en la reconstrucción.

Palabras clave: *Procesamiento de imágenes tridimensionales, Óptica de Fourier, Técnicas de fotografía computacional e imaginería, Técnicas de reconstrucción de imágenes*

Index

Figure Index

Table Index

Introduction

Computational optical imaging exploits the joint design of illumination, optics, detectors, and reconstruction algorithms to achieve system optimization suited for task-specific imaging, overcoming the limitations of conventional techniques [Lam, 2015]. It constitutes an interdisciplinary field including optics, image processing, and computer vision, among others [Zhou and Nayar, 2011, Levoy, 2010, Mendlovic, 2013, Lin et al., 2013, Lin et al., 2015] and its scope spans from the fundamental science to medical and entertainment applications.

The main idea behind *Computational optical imaging* approach consists on the acquisition of images through a modified optical system for its computational processing through algorithms, in order to give rise to new images with novel features that were originally absent in the acquired images.

Physics-based understanding and modeling of the underlying optical image formation process is of fundamental importance. In this regard, this thesis explores computational optical imaging methods based on the physics behind optical systems with limited depth-of-field. In this context, the originally acquired images are severely defocused, *i.e.* they have a shallow depth-of-field.

The manuscript is divided into seven chapters.

In Chapter 1, we will briefly see how the transition from analog cameras to digital cameras changed the role of photography in science. Its evolution from a simple documentation tool to a process involving a computer, that codifies image information which can be decoded with different purposes by post-processing algorithms, gave rise to *Computational optical imaging* methods, allowing to extend the functionality of traditional cameras.

The next chapter (Chapter 2) explores how physical limitations in acquisition optics led to images with a shallow depth-of-field. This shallow depth-of-field makes it impossible to acquire the whole object in-focus in a single image. In traditional mechanically based lens systems, the focal sweep or axial scanning of a three dimensional object is accomplished by relative movement between the optical elements within the optical system or by mechanically changing the relative distance between

the object and the optical system. However, in recent years a new kind of deformable lenses have been developed as an alternative to rigid lenses and are becoming an interesting option to the use of mechanical platforms. An electrically focus-tunable lens (ETL) allows to change the focal plane along the optical axis without mechanical movements. Finally, we introduce our experimental optical setup for image acquisition which uses an ETL and avoids registration between the acquired images [Alonso et al., 2015, Alonso, 2016a].

Chapter 3 addresses the problem of how to synthesize a single sharp image (an all-in-focus image) given a sequence of multi-focus images. We introduce the image formation model in Optics along with the concepts of the Impulse Response (or Point Spread Function, PSF) of an optical system and discuss whether it can be considered as depth-invariant or depth-variant. We mention different methods to accomplish image fusion and explain our proposed Fourier Domain method, showing its good performance in comparison with the state-of-the art [Alonso et al., 2015].

The purpose of Chapter 4 is to show that through the proposed Fourier Domain formalism, we are also capable of post-capture aperture reshaping and refocusing a 3D scene from a multi-focus stack to synthesize images as if they had been acquired with different pupils [Alonso et al., 2016, Alonso, 2016c]. Partially extended depth-of-field and all-in-focus image reconstruction are obtained as particular cases.

Free view-point image synthesis is treated in Chapter 5, where we present a method for post-capture perspective shifts reconstruction from a multi-focus image stack [Alonso et al., 2016]. Both horizontal and vertical directions are considered and comparisons with other methods are implemented. In the comparison, our method outperforms state-of-the-art schemes and unlike depth-invariant approaches, deals well with strongly defocused image stacks. It also works with no need for segmentation (extraction) of the focused regions or retrieval of the depth map, which could introduce inaccuracies into the reconstruction. Perspective shifts in axial direction are also addressed.

In Chapter 6, based on the view-point synthesis discussed in the previous chapter, we introduce a method to synthesize a pair of stereoscopic images of a 3D scene from a multi-focus image stack [Alonso, 2016b]. Reconstructed images are presented as stereoscopic pairs as well as red/cyan anaglyphs.

Finally, in Chapter 7, the conclusions and future lines of research are presented.

Chapter 1

Computational Optical Imaging

1.1 The advent of photography

Many scientists have been interested in color, light and vision along the time. Isaac Newton experimented with optics and gave an insight on colors of light. Through his experiments, he discovered that it was possible to combine all the colors together into white. His famous circular spectral arrangement, using seven colors[1] like the seven notes of a musical scale, appeared in 1704 in his *Opticks* (see Figure 1.1).

The trichromatic theory, which only need three primary colors to create white or any perceivable color, was first put forward by Thomas Young in 1802. He postulated the existence of three types of photoreceptors in the eye (now known as cone cells), each of which was sensitive to a particular range of visible light.
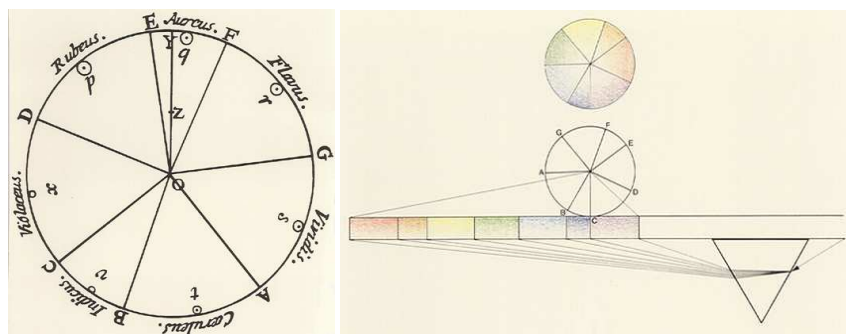
Figure 1.1: Newton's illustration of the additive color synthesis using seven colors like the seven notes of a musical scale (http://www.colorsystem.com/?page_id=683&lang=en)

In the meantime, the basic principle of photography was born in the nineteenth century with the invention of a light sensitive surface by Joseph Nicéphore Niépce.

---

[1] red, orange, yellow, green, blue, indigo and violet

Before that, the principle of the camera was known by means of the *camera obscura* (dark chamber). In the camera obscura, the light would pass through a small hole in the wall of a darkened box and project an upside down image of whatever was outside the box (see Figure 1.2). Niépce invention in 1826 allowed to register an image for the first time. In Figure 1.3 a camera obscura was used by Niépce to expose a copper plate coated in silver and pewter (it required about eight hours of exposition).
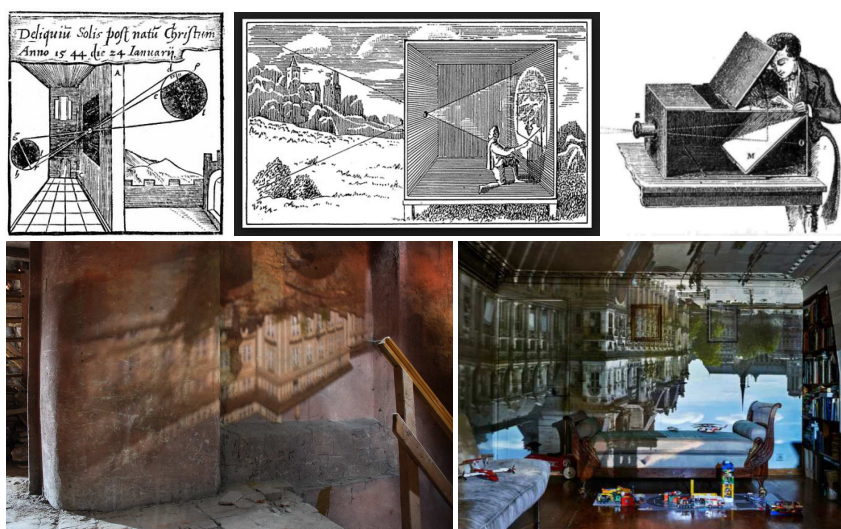
Figure 1.2: Camera obscura (pinhole camera). First row: principle of the camera obscura (from https://en.wikipedia.org/wiki/Camera_obscura); Second row: (left) image of the New Royal Palace at Prague Castle projected onto an attic wall by a hole in the tile roofing; and (right) Abelardo Morell, Camera obscura, View on Hotel de Ville, Paris, 2015 (from http://www.abelardomorell.net/project/camera-obscura/).

It was a few years later, that Louis Daguerre, who learned some of Niépce's techniques, announced to the French Academy of Sciences in Paris, in 1839, that he had improved a photographic imaging technology through a new process he named a *daguerreotype*. The daguerreotype was the first commercially successful photographic process in the history of photography. It significantly reduced exposure time and created a lasting result, but only produced a single image. Acquiring daguerreotypes was full of errors since plates had to be developed wherever the image was made (see Figure 1.4). At the same time, William Henry Fox Talbot was experimenting a new method, patented in February 1841. Talbot's innovations included the creation of a paper negative, and a new technology that involved the transformation of the negative to a positive image, allowing for more than one copy of the picture. Further

advances in technology continued to make photography more accessible (https://www.khanacademy.org/humanities/becoming-modern/early-photography/).

Figure 1.3: Joseph Nicéphore Niépce, *View from the Window at Gras* (1826) (https://www.khanacademy.org/humanities/becoming-modern/early-photography/).

Herman von Helmholtz [1] further developed Young's trichromacy theory in the 1850's, stating that the three types of cone photoreceptors could be classified as short-preferring (blue), middle-preferring (green), and long-preferring (red), according to their response to the wavelengths of light striking the retina. The relative strengths of the signals detected by the three types of cones are interpreted by the brain as a visible color (see Figure 1.5). His theory would then be refined by James Clerk Maxwell [Cat, 2013].

In 1861, Maxwell was invited to give a lecture at the Royal Institution on his work on color vision. Rather than talk about the principles, he preferred to give a visual demonstration that any color could be made by mixing the three primaries (red, green and blue: RGB). Thomas Sutton, a colleague at King's College London, was an expert photographer[2] and helped Maxwell. The basic techniques of black

---

[1] In 1851, Helmholtz revolutionized the field of ophthalmology with the invention of the ophthalmoscope; an instrument used to examine the inside of the human eye.

[2] Sutton was the inventor of the Single Lens Reflex camera (SLR) which he patented in 1861. Basically the same lens through which the light passes to expose the film is the same lens through which the photographer views the scene through the viewfinder. This design (with a sensor instead of film), is still popular nowadays under the name Digital Single Lens Reflex camera (DSRL).
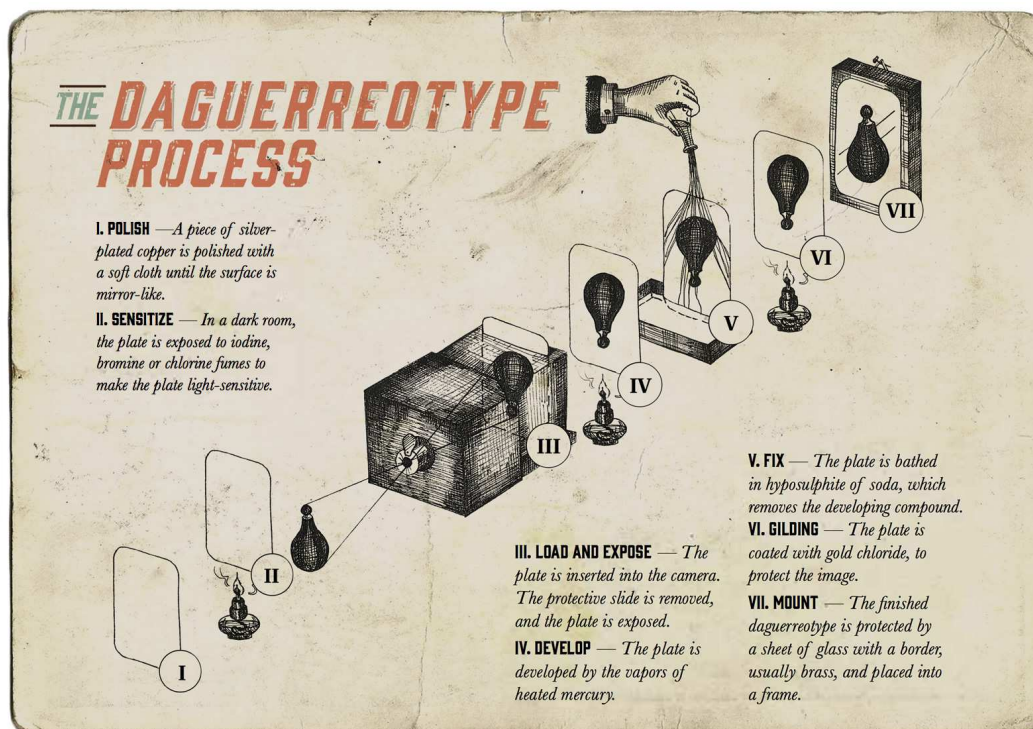
Figure 1.4: (Top) Daguerreotype process. (Illustration by Susanna Celeste Castelli, Densitydesign research lab; Source: George Eastman House International Museum of Photography and Film). (Bottom) Louis Daguerre, *The Artist's Studio*, 1837, daguerreotype. (https://www.khanacademy.org/humanities/becoming-modern/early-photography/)

and white photography were already known and it was already possible to project a photograph on a screen. They took three pictures of a tartan ribbon, through red, green and blue filters in turn (see Figure 1.6) and then projected them simultaneously on the screen using the same filters (see Figure 1.7). Combining the color projections back together, Maxwell and Sutton created the world's very first permanent color photograph which would become the basis for all color photography to come (http://filmmakeriq.com/lessons/the-history-and-science-of-color-film-from-isaac-newton-to-the-coen-brothers/).
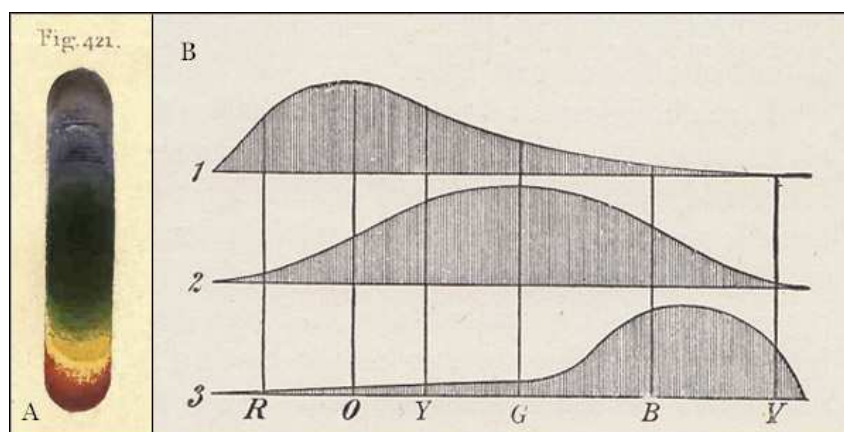
Figure 1.5: (Left) Thomas Young's diagram of the spectrum. (Right) Helmholtz's depiction of Young's theory. (http://www.huevaluechroma.com/062.php)

Figure 1.6: Red, Green and Blue images of the tartan ribbon acquired through the respective filters. (http://proyectoidis.org/james-clerk-maxwell/)

Around 1883 the American inventor Goerge Eastman, who found the Kodak Corporation, developed a flexible film that could be rolled, which made photography accessible to the general public. With many changes, the photographic roll
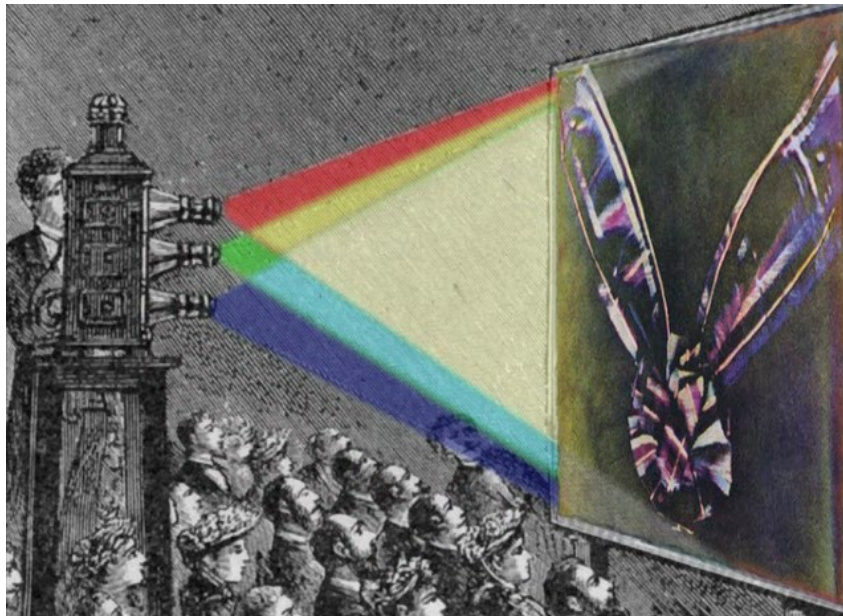
Figure 1.7: Illustration of additive color synthesis using three color projection (RGB). (http://proyectoidis.org/james-clerk-maxwell/)

film remained the dominant form of photography until the early 21st century, when advances in digital photography drew consumers to digital formats.

1.2 Rise of the digital era

Advents in semiconductor technology since the mid-twentieth century allowed for the emergence of two dominant types of digital sensors: the charge-coupled device (CCD), and the complementary metal-oxide semiconductor (CMOS) sensor. Both contain photoactive regions divided into a regular array of photodetectors, with each photodetector resulting in a pixel of the digital image. Consequently, there is no need to "develop" the picture from the film, nor is it necessary to purchase new recording materials (films) when one likes to take more pictures [Lam, 2011].

Photography is normally a non-invasive or non-intrusive technique and is commonly used for remote recording, sensing and measurement purposes. In particular, it is a means of extending imagery beyond the limits of human visual perception, producing permanent records for analysis and evaluation of the subject and processes involved. The role of photography in science started as a documentation tool for experimental results, *i.e.* to record the laboratory equipment and experimental setup. The advent of digital photo sensors along with the availability of ever faster
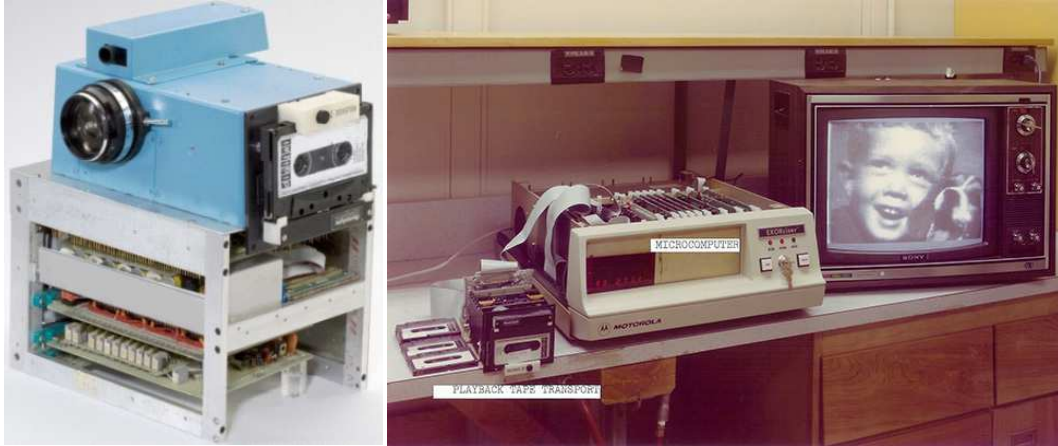
Figure 1.8: First digital camera, developed by Kodak engineer Steve Sasson in 1975 and patented in 1978. Left: acquisition, and right: display (from http://petapixel.com/2010/08/05/the-worlds-first-digital-camera-by-kodak-and-steve-sasson/).

computer technology has given rise to the establishment of new fields of research such as digital image processing, computer vision, and computational photography.

There is an ongoing demand on behalf of the scientific and consumer industries to make lighter, higher resolution cameras with extended functionalities controlled by algorithms. This has brought attention to finding new ways to develop hardware modifications in a camera that meet more demanding system's requirements. In this sense, the sole documentation character of a camera has evolved into a codification tool susceptible to hardware modifications (for example by changing optical components) that relies on modeling and computational methods in order to provide a final image with valuable scene information [Lam, 2011]. To adapt a digital camera to accomplish a particular task requires understanding its capabilities and contemplate the trade-offs between performance and cost [Ray, 2015].

1.3 Computational optical imaging

Computational photography and imaging techniques are rapidly evolving as an interdisciplinary research field known as *Computational optical imaging*, including optics, image processing, and computer vision, among others [Zhou and Nayar, 2011, Levoy, 2010, Mendlovic, 2013, Lin et al., 2013, Lin et al., 2015]. Computational optical imaging exploits the joint design of illumination, optics, detectors, and reconstruction algorithms to achieve a system optimally suited for task-specific imaging, overcoming the limitations of conventional techniques [Lam, 2015]. Its scope spans from

the fundamental science to medical and entertainment applications. Most of the reconstruction algorithms often aim at solving an ill-posed, linear inverse problem [Lam, 2011].

An example of the aforementioned techniques is the use of aperture patterns for defocus deblurring. When diffraction and optical aberrations are negligible, the shape of the point-spread function (PSF) is simply determined by the aperture pattern, and the scale is determined by the amount of defocus. In order to synthesize a specific PSF, *e.g.* for defocus deblurring or for obtaining a depth-map by means of depth-from-defocus techniques, the shape of the aperture has to be designed through a *coded aperture* (see Figure 1.9)[Zhou et al., 2011].

Optical imaging acquisition combined with post-capture processing algorithms also enables the reconstruction of images with novel features of interest such as an extended depth-of-field (all in focus), novel viewpoints of a scene, refocusing, etc. A set of multiple images (taken at once through a microlens or lenslet array, or in a time sequence) often encloses the three-dimensional (3D) information of an underlying 3D object, and a particular model along with computational processing are needed in order to recover this information.

Plenoptic cameras, also known as *light field* cameras, have been developed with the idea that they can collect visual information that is irreversibly lost in one-shot conventional camera acquisition. The plenoptic camera makes use of a conventional main lens, but in addition, places a lenticular array in the original camera sensor position Figure 1.10.

Light field photography is a ray model approach where a modified camera gathers information about the ray direction by means of a lenslet array inserted in front of the camera sensor [Ng et al., 2005, Mendlovic, 2013]. The light field camera approach implies the simultaneous acquisition of several views on a single image, but this instant capture of spatially multiplexed information of a scene comes at the cost of a reduction in the spatial resolution of the reconstructed images (one pixel per microlens) [Levoy, 2006, Lam, 2015]. Another approach to the spatially multiplexed acquisition of information from a scene is given by *integral imaging*, in which the lenslet array is positioned in the main lens plane [Park et al., 2008, Xiao et al., 2013, Martínez-Corral et al., 2014].

Plenoptic cameras are also able to reconstruct (low resolution) images resembling different camera positions by synthesizing pin-hole images rather than full-aperture images, obtaining changes in parallax or perspective [Ng, 2006]. This capability can be used for rendering stereoscopic views [Huang et al., 2015] which are of interest for many areas ranging from medical imaging [van Beurden et al., 2012] to visual entertainment applications [Mendiburu, 2012].
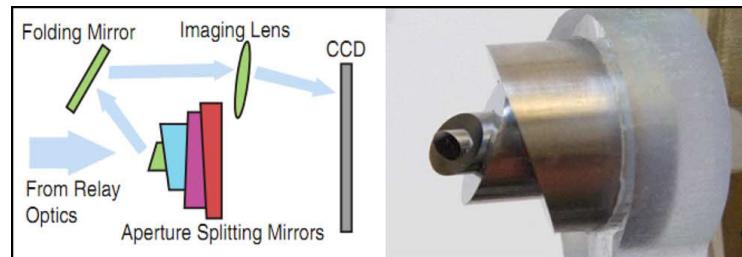
(a) Multi-aperture photography using tilted mirrors
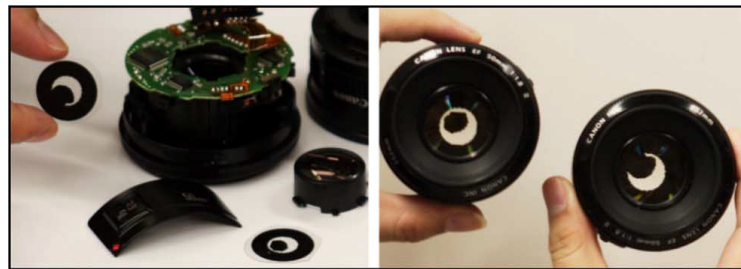
(b) Coded aperture using photomasks

A Scroll of Paper Patterns   Liquid Crystal Array

3cm
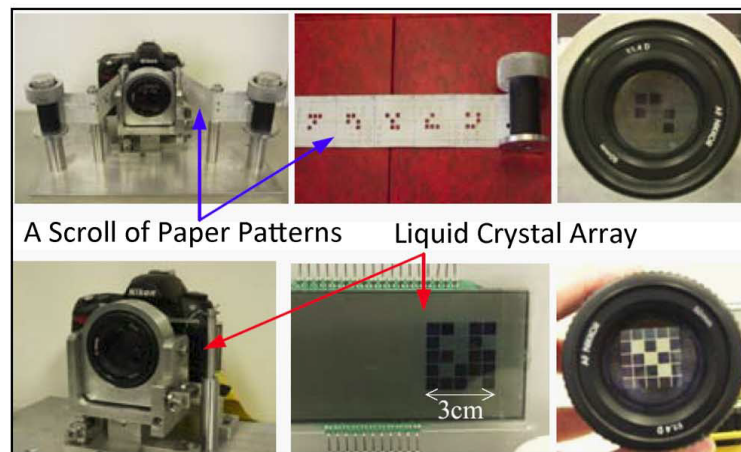
(c) Programmable aperture camera using paper patterns or LCD

Figure 1.9: Coded aperture implementations. (a) Use of a four-way aperture-splitting mirror to divide the lens aperture; (b) photomasks inserted into lenses; (c) scroll of paper patterns and LCD pattern. From [Zhou and Nayar, 2011] and references therein.

11
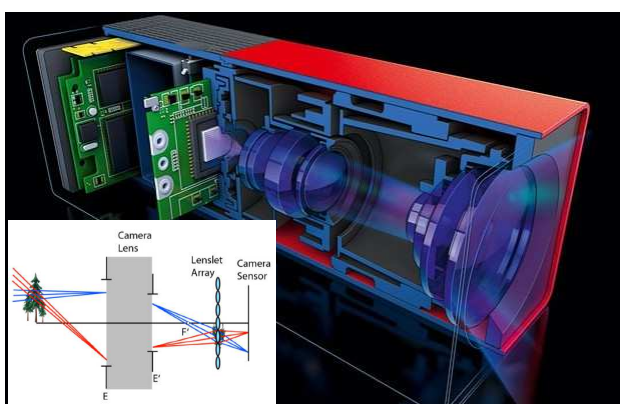
Figure 1.10: (Top) Captured image on the sensor through the microlens array. (Bottom, Left) Interior of the Lytro (comercial system under this principle) (from [Ng et al., 2005]). Inset with a scheme of the optical system (from [Schwiegerling and Tyo, 2013]). (Bottom, Right) The low resolution extended-depth-of-field (EDOF) rendered image (from [Ng et al., 2005]).
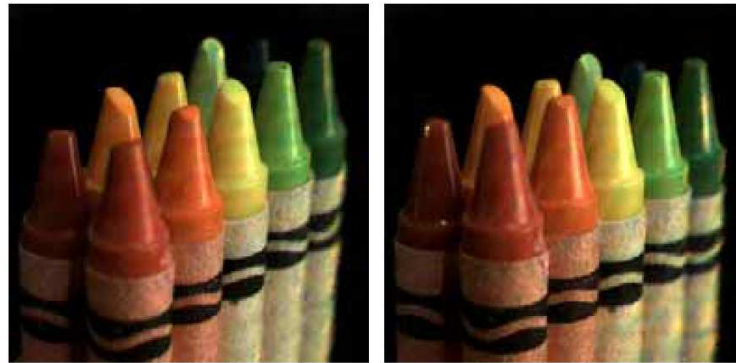
(A): High viewpoint   (B): Low viewpoint

Figure 1.11: Changes in perspective by synthetically moving the camera (A) up and (B) down (from [Ng, 2006]).

The ability to handle the optical settings, such as focus and viewpoint, by means of a plenoptic camera can also be accomplished by a conventional camera if the information of the scene is acquired in a time-sequential capture of a multi-focus image stack (*i.e.* temporally multiplexed information with a high spatial resolution instead of one shot capturing several low spatial resolution views). Even though the temporally multiplexed approach comes at the cost of losing the instant capture of the scene, the development of new generations of *electrically focus-tunable lenses* (ETLs) with increasingly fast responses [Lu and Hua, 2015] that allow high-speed z-axis scanning of the scene is bridging the gap between high spatial resolution and instant capture.

Chapter 2

Multi-focus image acquisition

In this chapter we will briefly review geometrical image formation through a thin lens and the limited depth-of-field (DOF) due to the finite aperture of the lens given by its own edge or a diaphragm in the plane of the lens. Then we will see how the conventional strategies to acquire a multi-focus image set (multiple differently focused images) of a 3D scene lead to differences in scale or alignment between the acquired images. Sometimes these differences may be just ignored but in many situations a pre-processing procedure called image registration needs to be performed. However, as image registration is computationally expensive it is desirable to avoid it when possible. We then propose a simple optical system in order to acquire real multi-focus images which does not require image registration pre-processing. Besides, for continuous 3D scenes (scenes that contain volumetric objects), we develop a recurrence formula based on DOF to optimize the number of images to be acquired in order to perform a discrete focal sweep along the optical axis.

2.1   Image formation by an ideal thin lens

Geometric optics approaches light as rays that are refracted and reflected, according to reflection and refraction laws, at places where the indices of refraction change. A *thin* lens is that whose thickness is considered small in comparison with the distances generally associated with its optical properties (radii of curvature of the two spherical surfaces, primary and secondary focal lengths, and object and image distances) [Jenkins and White, 1957]. If an object is placed in front of a converging lens, beyond the front focal plane, then under appropriate conditions there will appear across a second plane a distribution of light intensity that closely resembles the object. This distribution of intensity is called an image of the object. All light rays that originate at some specific point on the object and then go through the lens are

redirected to converge at a single point on the image [Born and Wolf, 2000, Goodman, 1996]. The condition for image formation is given by the well known thin lens law in Equation (2.1) (in its Gaussian form):

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{fl} \tag{2.1}$$

where d_o is the distance from the object to the lens, d_i is the distance from the lens to the image, and fl is the focal length of the lens (the focal length of the lens is the distance from the lens at which rays parallel to the optical axis converge). Figure 2.1 shows the geometry of image formation by an ideal thin lens. All rays from an object
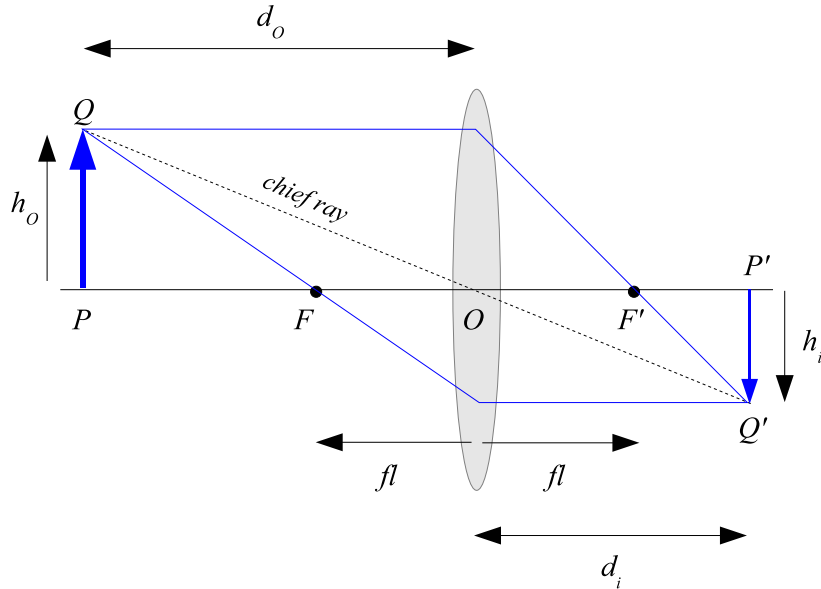
Figure 2.1: Image formation through a thin lens.

point Q, which pass through the lens, are refracted to pass through the image point Q' and the rays from another point P are refracted to pass through P'. Any ray that passes through the center of the lens is undeviated (*i.e.* it experiences no refraction) and is referred as chief ray[1]. If the object is moved closer to the primary focal point, F, the image will be formed farther away from the rear or secondary focal point, F', and will be larger, *i.e.* magnified. If the object is moved farther away from F, the image will be formed closer to F' and will be smaller in size. Such ideal conditions hold only for paraxial rays, *i.e.* rays close to and making small angles with the lens

---

[1]For systems with entrance and exit pupils which do not coincide with the lens, the chief ray passes through the center of the entrance and after suffering refraction in the lens passes through the center of the exit pupil [Jenkins and White, 1957]. However in our approach we will only consider the case in which the pupil is in the same plane as the lens.

axis.

The image magnification produced by a single lens may be derived from the geometry of Figure 2.1. By construction it is seen that the triangles QPO and $Q'P'O$ are similar. Corresponding sides are therefore proportional to each other, so that

$$\frac{P'Q'}{PQ} = \frac{OP'}{OP} = \frac{d_i}{d_o} \tag{2.2}$$

Taking upward directions as positive, $h_o = PQ$, and $h_i = -Q'P'$. The *lateral magnification* is therefore

$$M = \frac{h_i}{h_o} = -\frac{d_i}{d_o} \tag{2.3}$$

When d_i and d_o are both positive, as in Figure 2.1, the negative sign of the magnification means an inverted image.

2.1.1 Depth of field

In the image formed by a convex lens (or an optical system), objects at a particular distance from the lens will be focused whereas objects at other distances will be blurred or defocused by varying degrees depending on their distances. Only those light rays that come from the points of the object which are at the focusing distance will be converging correctly onto the image plane and then will be imaged in-focus. Any light rays coming from objects at different distances from the lens are going to either converge before or after the image plane creating a blurry spot at the image plane [Subbarao and Surya, 1994].

The depth of field (DOF) is the range of scene depths that appear focused in an image for a given focal length and aperture of the optical system. Conventional cameras are limited by a fundamental trade-off between depth of field and signal-to-noise ratio (SNR). DOF can be increased by making the aperture smaller, however, this reduces the amount of light received by the detector, resulting in greater image noise (lower SNR). This trade-off gets worse with increase in spatial resolution (decrease in pixel size). As pixels get smaller, DOF decreases since the defocus blur occupies a greater number of pixels (this is related to the concept of circle of confusion that we will be treated in Section 2.4). At the same time, each pixel receives less light and hence SNR falls as well [Kuthirummal et al., 2011].

DOF changes along the optical axis being shallower when focusing objects near the camera and deeper when focusing objects far from it (see Figure 2.3). We will come back to DOF in Section 2.4.

Figure 2.2: Limited DOF in optical systems. *a*) illustration of the concept of shallow DOF (CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=330435), *b*) DOF in everyday photography (J.Alonso), *c*) DOF in macro-photography (digital-photography-school.com) and *d*) DOF in photomicrography (image taken through a microscope). (Wim van Egmond http://www.nikonsmallworld.com/galleries/photo/2013-photomicrography-competition)
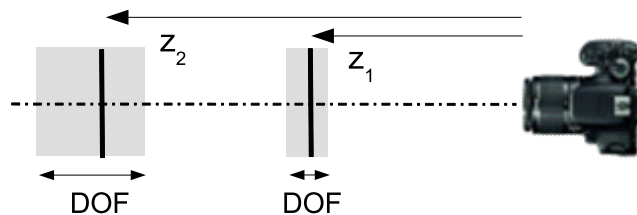
Figure 2.3: Along the optical axis, two focal planes z_1 and z_2 (full vertical lines) have different DOF's. DOF is shallower for nearest focusing distances while becomes larger for farther focusing distances.

18

2.2 Moving the camera or the object to acquire the multi-focus image stack

Conventional image fusion techniques of multiple differently focused images demand a registration between the acquired images in the preprocessing step. The acquired images may differ in displacements, scale and rotation due to the change of the focal plane or the position of the camera. Then, the differently focused images need to be corrected in such a way that they have the same scale and the same field of view.

Image registration is one of the most important image processing applications of geometric transformations [Gonzalez et al., 2009] and a key issue in image fusion as misalignment produces severe edge artifacts in the combined images [Pajares and De La Cruz, 2004]. In particular, in multi-focus image fusion, due to the change of focus settings or movements of the object or the camera, the images may undergo a scale change that must be corrected during preprocessing [Subbarao and Surya, 1994, Nayar and Nakagawa, 1994, Schechner et al., 2000]. For example, in microscopy, 3D images (*i.e.* images from a thick specimen) are assembled from a focal stack of 2D images acquired using sequential refocusing. The set of images is usually acquired by mechanical movement of the microscope stage holding the sample or by moving the objective [Abrahamsson et al., 2013]. The problem is illustrated in Figure 2.4.

Normalization with respect to image magnification is complicated. It can be done by image interpolation and resampling such that all images correspond to the same field of view [Subbarao and Surya, 1994] or by means of hierarchical block matching in which displacement, scale and rotation are taken into account [Kubota et al., 1999].

For example in [Kubota et al., 2007] the input images are taken for the same camera position with different foci on two depths (foreground and background). As the change in focusing distance is implemented through an internal relative displacement of the many lenses that constitute the optics of the camera, a pre-processing image registration step needs to be implemented to correct the difference of magnification due to focusing. The sensor can also be displaced, like in [Kuthirummal et al., 2011] where the acquisition optical system is adapted to change the position of the sensor relative to the lens and it is pointed out that when the detector is translated, the magnification of the imaging system changes. The result is that the images are not correct in perspective, but as the changes in magnification for their working distances are very small, the distortions are also small enough to be ignored. In fact in many practical applications involving thin samples the magnification change can be ignored [Subbarao and Surya, 1994]. Clearly this is not the case for extended axial objects like thick samples in microscopy.

19
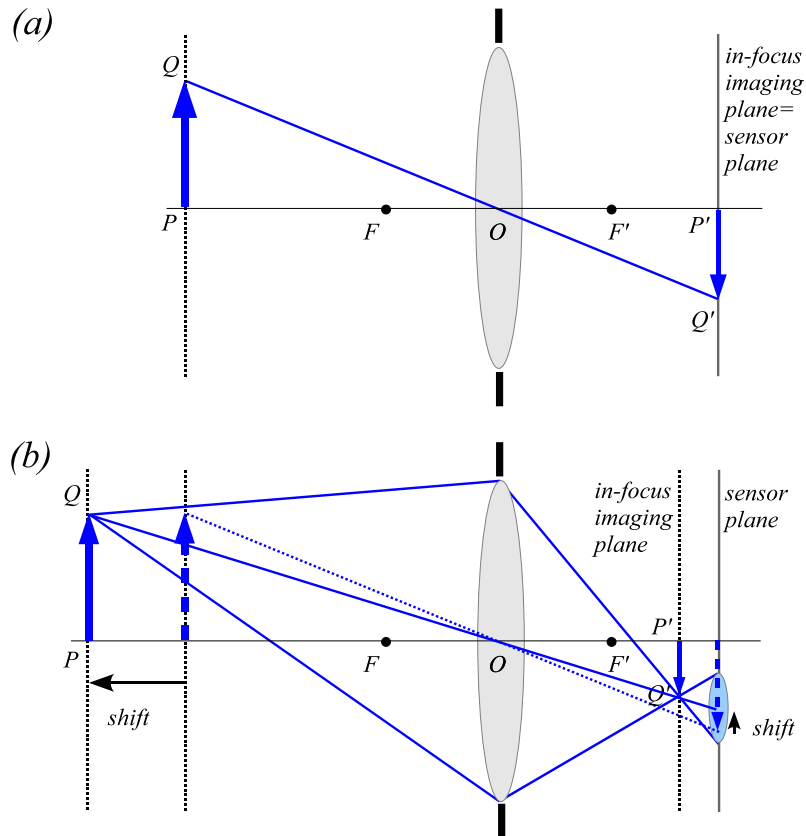
Figure 2.4: (a) in-focus image of the tip Q of the arrow PQ through a fixed focal length lens; (b)out-of-focus image of Q due to a relative displacement between the object and the lens. The sensor plane is no longer the imaging plane and the point Q is imaged as a uniform circle in the camera. Note that the chief ray changes between (a) and (b) implying a change in scale between the captured images (a) and (b). There is a shift between the position of the tip of the arrow when imaged in-focus at the sensor (dashed arrow) and the center of the circle when imaged out-of-focus.

Registration is computationally expensive even by using hierarchical block matching [Kubota et al., 1999, Je and Park, 2013], so it is desirable to avoid it whenever possible. An alternative is to work with telecentric systems. These systems provide a nearly constant magnification over a range of working distances, which means that object motion towards or away from the lens does not affect image magnification *i.e.* it does not change the image size [Watanabe and Nayar, 1995, Martínez-Corral et al., 2015b]. Telecentric systems eliminate perspective, which may be desirable for some applications but not for others. Besides, telecentric designs tend to be bigger and more complex than non-telecentric designs [1].

2.3 Proposed setup for image acquisition

In the present section we will describe the optical system that we have implemented in order to acquire real multi-focus images with shallow DOF from 3D-scenes. We will show that the proposed system does not require image registration [Alonso et al., 2015, Alonso, 2016a].

2.3.1 Electrically focus-tunable lenses

In traditional mechanically based lens systems, the focal sweep or axial scanning is accomplished by the relative motion between the optical elements within the optical system or by mechanically changing the relative distance between the object and the optical system. However, in recent years a new kind of deformable lenses have been developed as an alternative to the use of rigid lenses and mechanical platforms. Basically, an electrically focus-tunable lens (ETL) is a lens whose focus can be tuned electrically for performing non-mechanical axial scanning.

We will consider the Optotune EL10-30 with focal distance in the range $55 - 165mm$. The core element of this shape-changing lens consists of a thin membrane that builds an interface between a liquid-filled chamber and air. The EL10-30 has an electromagnetic actuator that is used to exert pressure on the liquid reservoir surrounding the clear aperture of the lens. Fluid is thus forced into the center of the lens, changing the curvature of the membrane. Hence, the focal distance of the lens is controlled by the current flowing through the coil of the actuator (see Figure 2.5). Controlling the electrically tunable lens (ETL) requires a standard power supply as a current source (lens driver) to supply the lens with a current between 0 and 280mA as indicated by the manufacturer (http://www.optotune.com/).

---

[1] http://www.edmundoptics.cn/downloads/telecentricity.pdf
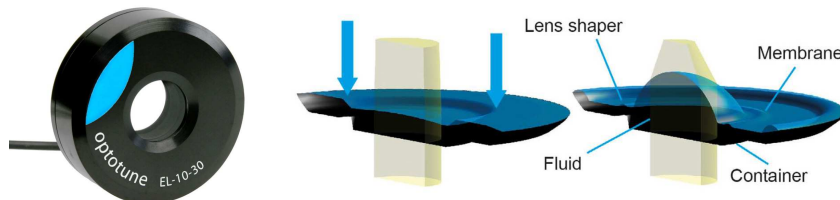
Figure 2.5: Working principle of the ETL Optotune EL-10-30.
http://www.optotune.com/

2.3.2 Calibration of the ETL

We performed a calibration of the lens following the procedure depicted in Figure 2.6.
A white LED (as an approximation to a point light source when a pinhole is placed
in front of it) is placed in the focal point of a lens with fixed focal length to generate
a collimated beam and the ETL is placed after it. Under this configuration, the
light converges on the focal point of the ETL, which can be varied according to
the applied current. Measuring the focusing distances ($\pm 1mm$) for different current
values allows one to obtain a calibration curve. Focal distances varying in $10mm$
steps were considered and the values of the corresponding current intensities were
gathered. The focal length of the ETL decreases with increasing current. From the
acquired data, for a given current it is possible to interpolate the corresponding focal
length and viceversa.

2.3.3 ETL-camera system

An object placed in the focal plane of the ETL is to be imaged at infinity through
that lens. Then, we attach the ETL right in front of the camera with its own lens
focusing at infinity (*i.e.* the focal length of the camera equals the distance d from
its lens to the CCD sensor). Under this configuration an object placed at the focal
plane of the ETL will be imaged (in-focus) at the camera's sensor plane. We think
of the ETL and the camera as two thin lenses placed in contact. This combination
behaves as a single thin lens with an effective focal length given by [Jenkins and
White, 1957]:

$$\frac{1}{fl_{eff}} = \frac{1}{fl_1} + \frac{1}{fl_2} \tag{2.4}$$

where fl_1 is the focal length of the ETL and fl_2 is the focal length of the camera's
own lens.

Figure 2.7 shows a scheme for this optical system. Consider an object placed
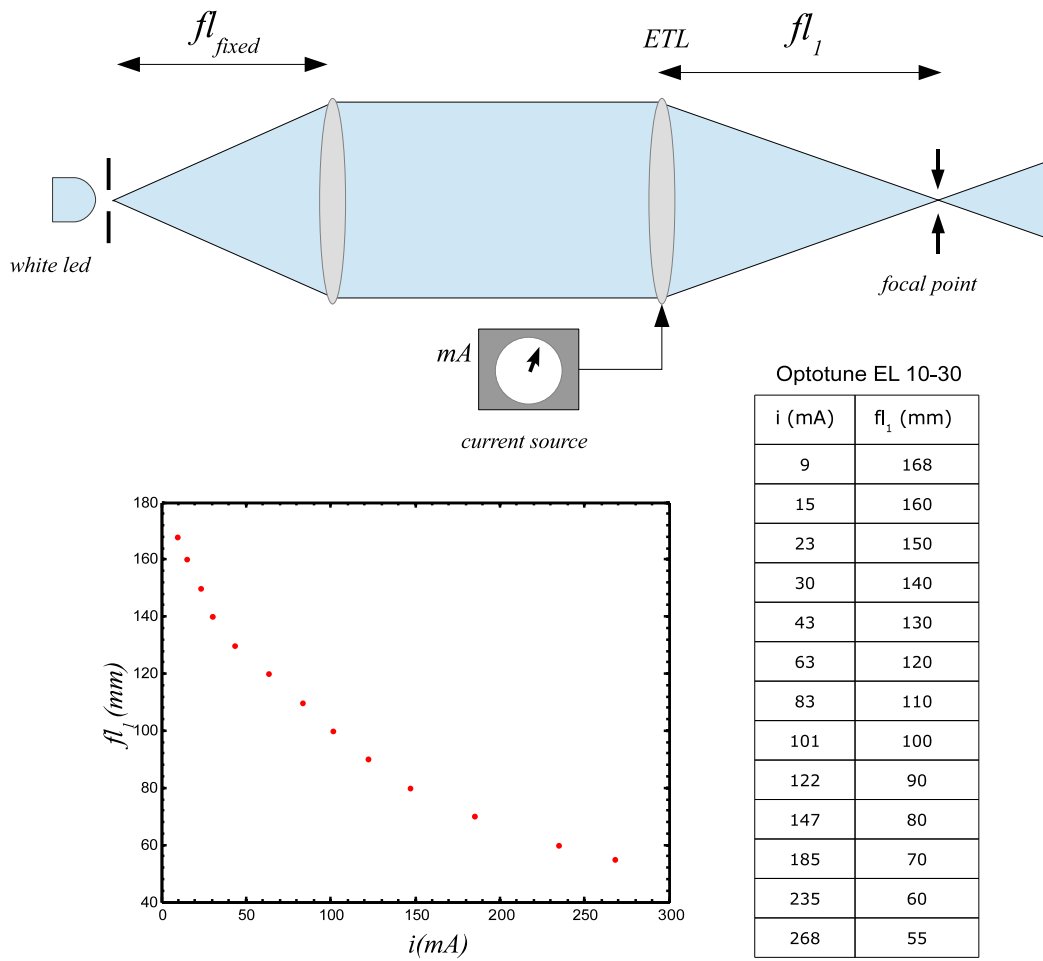
Figure 2.6: Calibration of the ETL. Scheme of the implemented procedure and data collected.

Optotune EL 10-30

i (mA)	fl_1 (mm)
9	168
15	160
23	150
30	140
43	130
63	120
83	110
101	100
122	90
147	80
185	70
235	60
268	55

at a distance z_0. If we set the focal length of the ETL to be $fl_1 = z_0$ then, as the camera's own lens is set to infinity ($fl_2 = d$), the image plane corresponding to that object coincides with the sensor plane and the object will be seen in-focus at the sensor plane. Thus, a point in the object will be seen as a point in a captured image. In this case, the total magnification of the in-focus image will be given by:

$$M_T = -\frac{d_i}{d_0} = -\frac{fl_2}{fl_1} = -\frac{d}{z_0}. \tag{2.5}$$

Objects will be imaged smaller and inverted in the sensor of the camera. Also closer objects will be imaged bigger at the camera sensor than farther objects.

On the other hand, if we tune the ETL in a way that $fl_1 \neq z_0$, then the image plane of the object will no longer be at the sensor plane and will appear out-of-focus in a captured image. In this case, a point in the object will no longer be seen as a point in the sensor plane but as a circle with uniform intensity (sometimes called the blur circle). Every point in the object is imaged as a blur circle whose center is at the same position at the sensor plane as it is the imaged point of the object when imaged in-focus at the sensor. Note that the *chief ray*[1] for a given point in the object stays the same regardless of the value of the focal length of the ETL (since the lenses that form our system are close to each other). Then, there is no need for registration (scale normalization or alignment) between the images acquired through this system. The principle of concept of the proposed optical system is illustrated in Figure 2.8.

We used a DCU224C Thorlabs color digital camera with 1280×1024 pixels ($5.65\mu m \times 5.65\mu m$ pixel size), with its lens MVL16L-16mm. The doublet is formed by the ETL directly attached in front of the camera lens as shown in Figure 2.9. For a given scene, the set of multi-focus images is acquired at the same exposure time. As an example of acquired real images, Figure 2.10 shows a set of images corresponding to four colored push pins placed at different distances from the optical system.

2.4 Optimal multi-focus image sequence

Image acquisition is a two dimensional projection over the camera's sensor of a three-dimensional object or scene. For plane objects (*i.e.* objects in a plane perpendicular to the optical axis) this is not a problem. On the other hand, imaging objects whose intensity distribution may be a continuous distribution in space is a more complex situation but we can think of the 3D-scene as made of successive layers.

In this sense, a continuous 3D scene distribution can always be completely sam-

---

[1]ray passing through the center of the entrance aperture
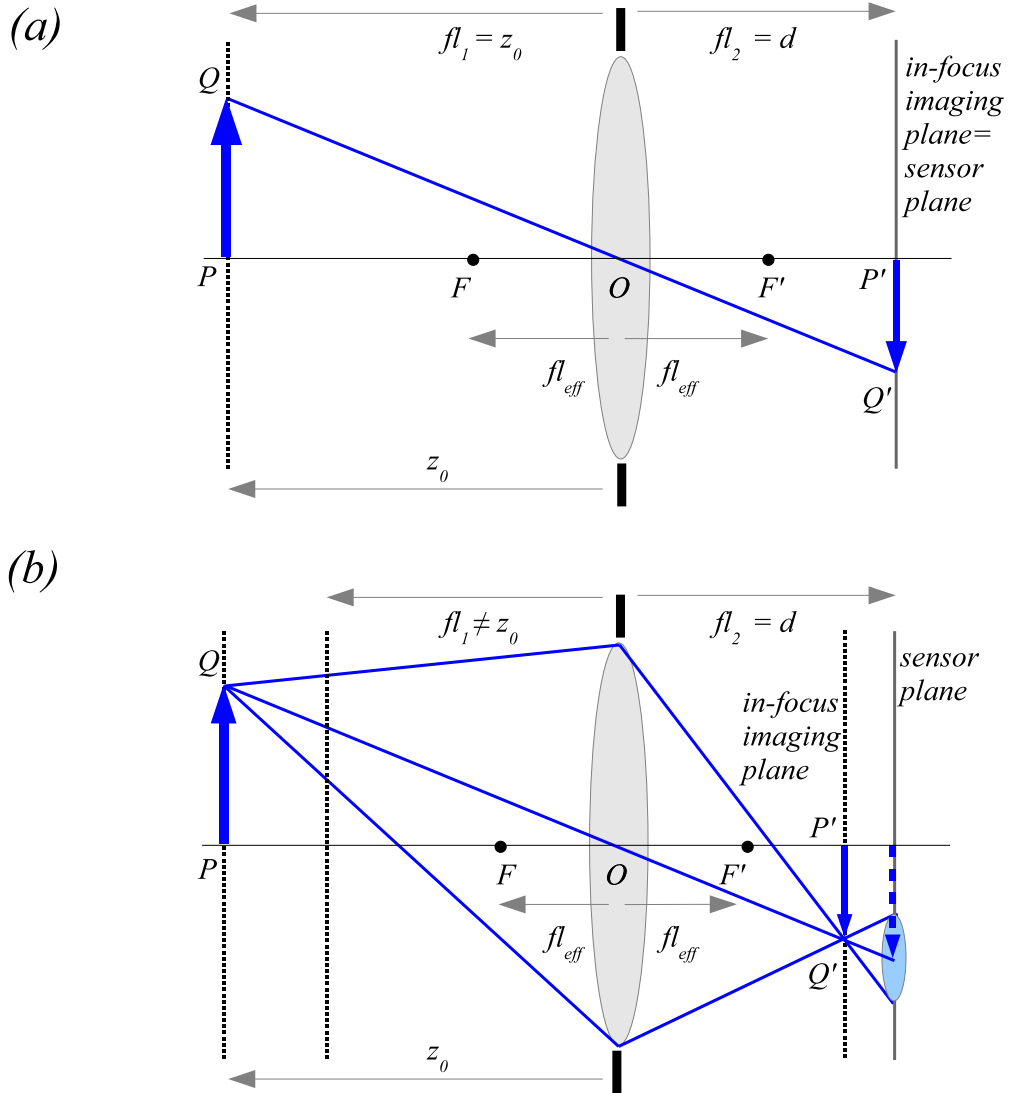
Figure 2.7: (a) Optical system with the ETL focusing at z_0. The object at z_0 is imaged in-focus on the sensor plane; (b) For a different value for the focal length fl_1 of the ETL ($fl_1 \neq z_0$), the in-focus image plane will not be at the sensor and then the object will be imaged out-of-focus at the sensor plane. Each point in the object is imaged as a blur circle whose center coincides with the in-focus image (*i.e.* for $fl_1 = z_0$) of the same point at the sensor.
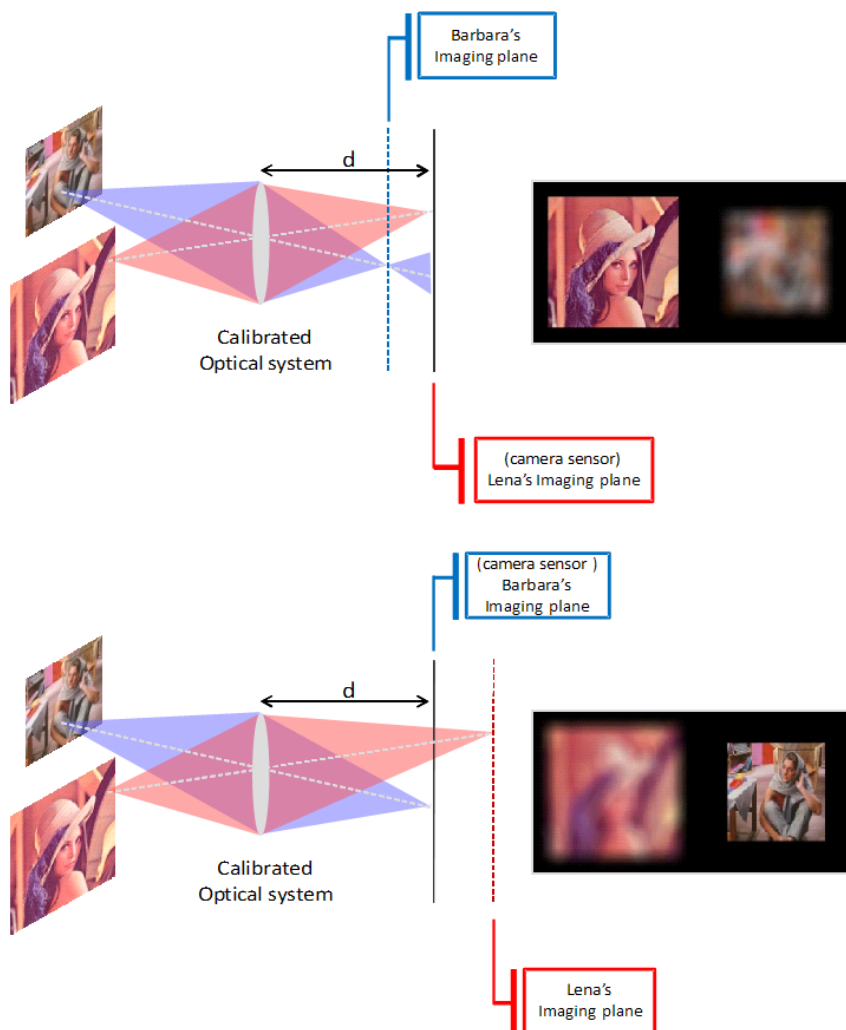
Figure 2.8: Principle of concept of the proposed optical system. Lena is the woman with the hat while the other is Barbara. Above, is represented the optical system focusing on Lena while below it is focusing on Barbara. On the right hand the simulated images as they will be captured in the camera are shown.
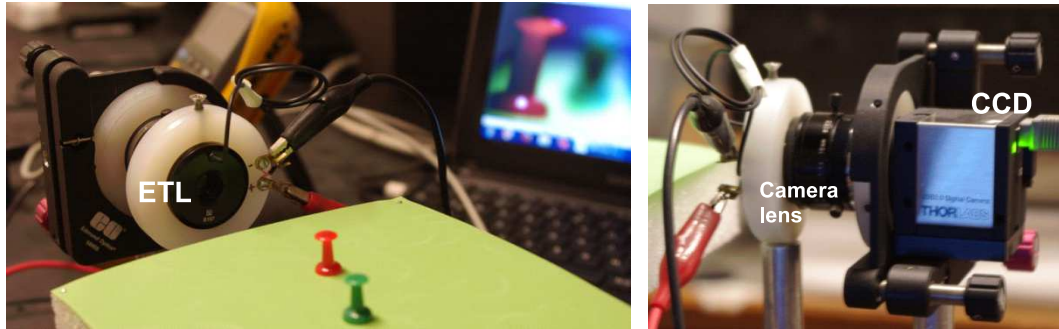
Figure 2.9: Optical system setup. (left) Front view of the setup where the ETL can be clearly seen; (right) Lateral view of the setup with the ETL directly attached to the CCD camera's lens.
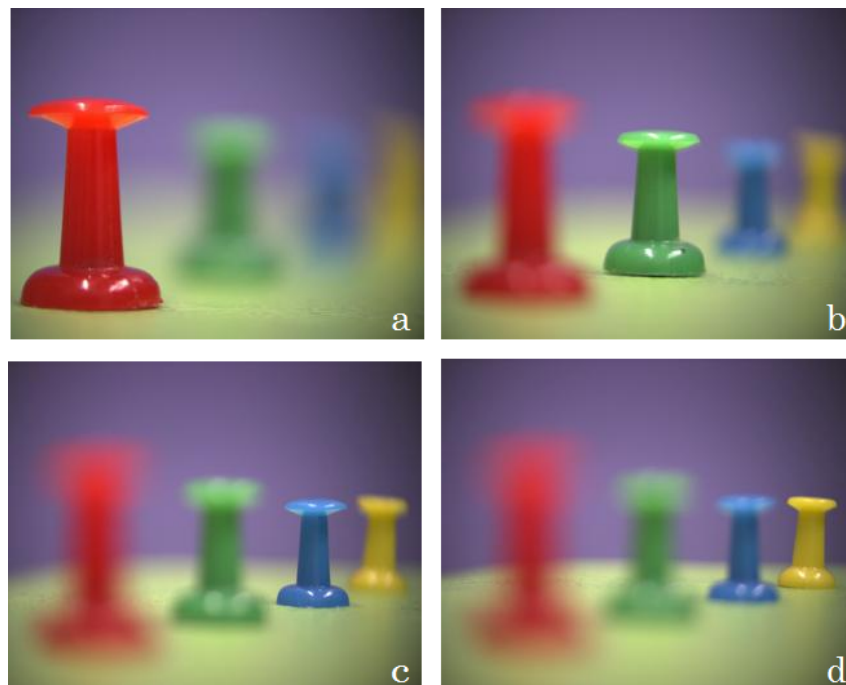
Figure 2.10: Real images of colored push pins captured with the system focused at $fl_1 =$ (a) $69mm$ (red), (b) $99mm$ (green), (c)$129mm$ (blue), and (d) $159mm$ (yellow).

pled in depth to form a z-stack, provided the discrete sequence of focusing distances is conveniently selected. We will see how we can use on our favor the shallow DOF to calculate an optimal sequence of focal distances for imaging the 3D-scene in a piecewise-planar approximation. The main idea is to break the 3D-scene in a number of N multi-focus 2D-planes without gaps or overlaps between DOFs from consecutive focal planes [Alonso et al., 2016]. A similar approach for different optical setups can be found in [Zhou et al., 2012].

For a given pupil (e.g., the circular aperture) in the lens plane of a camera and a given scene, we are interested in finding the optimal sequence of focusing distances in order to cover the scene in depth without gaps or overlaps between focused regions in the different images of the z-stack. A more detailed sweep will unnecessarily result in the same part of the object in focus in more than one image, implying more images to process and then demanding more computational resources. In this way it is possible to use a minimal set of images taking into account the dependence of DOF on the depth (unlike 3D deconvolution, which requires equally spaced focal planes in z [Sarder and Nehorai, 2006]).

Let us consider the optical system focused at distance z_k from the lens (Figure 2.11). There is a distance range around z_k that we will note as DOF_k (depth of field associated with z_k) for which the scene appears to be in focus in the imaging plane. This is due to the fact that, given the discrete nature of the sensor, there is the small amount of defocus which can be tolerated (*i.e.* non detected by the sensor). The *circle of confusion* (COC) is defined as the largest blur spot for which a point still appears as a single point (in focus) in an image.

In order to determine the optimal set of focal distances let us find the near $(z_k^{(1)})$ and far $(z_k^{(2)})$ limit of DOF_k (Figure 2.11). For an object in the near/far limit there is a corresponding perfectly in-focus image at $u_k^{(1)}/u_k^{(2)}$ which is related to $z_k^{(1)}/z_k^{(2)}$ through the Gaussian lens law (Equation (2.1))

$$\frac{1}{z_k^{(1)}} + \frac{1}{u_k^{(1)}} = \frac{1}{fl} \tag{2.6a}$$

$$\frac{1}{z_k^{(2)}} + \frac{1}{u_k^{(2)}} = \frac{1}{fl} \tag{2.6b}$$

where fl is the effective focal length for the system focusing at z_k. Since the same law holds for d (fixed distance between the imaging plane and the lens) and z_k:

$$\frac{1}{z_k} + \frac{1}{d} = \frac{1}{fl} \quad , \tag{2.7}$$
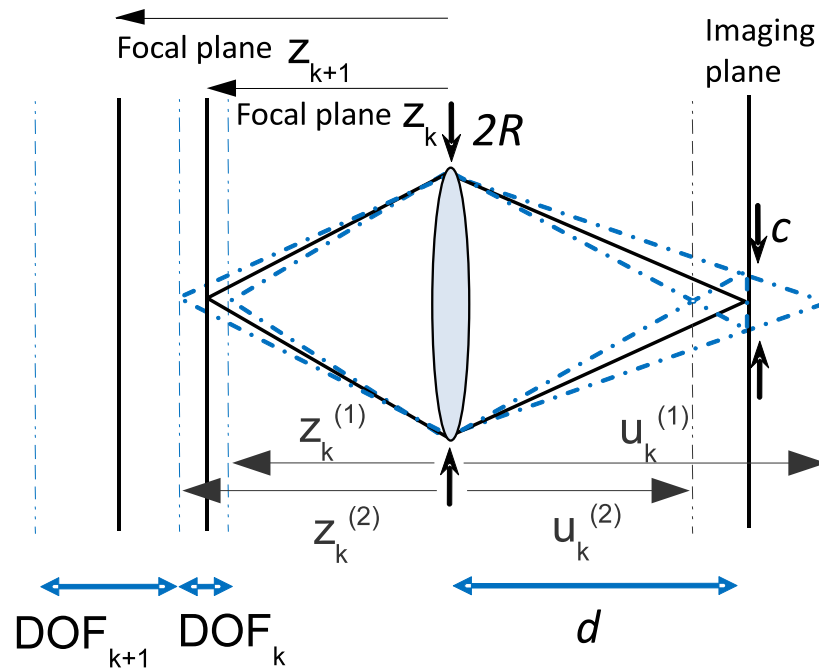
Figure 2.11: Optimal sequence of focusing distances for the given optical system: the 3D scene can be sampled along the optical axis in a way that two consecutive focal planes k and $k+1$ (full vertical lines on the left) provide adjacent and non-overlapping DOF's.

we can eliminate $\frac{1}{fl}$ between Equations (2.6a) and (2.7) and between Equations (2.6b) and (2.7) so the previous system simplifies to

$$\frac{1}{z_k^{(1)}} + \frac{1}{u_k^{(1)}} = \frac{1}{z_k} + \frac{1}{d} \tag{2.8a}$$

$$\frac{1}{z_k^{(2)}} + \frac{1}{u_k^{(2)}} = \frac{1}{z_k} + \frac{1}{d} \quad . \tag{2.8b}$$

From triangular similarity in Figure 2.11, it is easy to verify that the following relations hold:

$$\frac{c}{2R} = \frac{u_k^{(1)} - d}{u_k^{(1)}} \tag{2.9a}$$

$$\frac{c}{2R} = \frac{d - u_k^{(2)}}{u_k^{(2)}}, \tag{2.9b}$$

where R is the radius of the pupil aperture and c is the diameter of the circle of confusion (typically two pixels in size, assuming negligible diffraction[1]). The previous Equations can be rewritten as:

$$\frac{c}{2Rd} = \frac{1}{d} - \frac{1}{u_k^{(1)}} \tag{2.10a}$$

$$\frac{c}{2Rd} = \frac{1}{u_k^{(2)}} - \frac{1}{d}. \tag{2.10b}$$

By combining Equations (2.8a) and (2.10a) and combining Equations (2.8b) and (2.10b) we arrive at the following expressions for $z_k^{(1)}$ and $z_k^{(2)}$:

$$\frac{1}{z_k^{(1)}} - \frac{1}{z_k} = \frac{c}{2Rd} \quad , \tag{2.11a}$$

$$\frac{1}{z_k} - \frac{1}{z_k^{(2)}} = \frac{c}{2Rd}. \tag{2.11b}$$

---

[1]Note: The radius of the Airy function due to the limited circular aperture of the lens is given by $r_{Airy} \approx 0.61 \frac{\lambda d}{R}$ [Goodman, 1996, Iizuka, 2013], in order to neglect diffraction it should be smaller than the pixel size p. For the considered optical system $2R = 13.5mm$, $d = 16mm$, $\lambda = 550nm$ and $p \approx 5.65\mu m$, so $\frac{r_{Airy}}{p} \approx 0.14$ and therefore diffraction effects do not need to be considered.

The condition to obtain an optimal image stack is that the far limit of DOF_k coincides exactly with the near limit of DOF_{k+1} of the next focal distance z_{k+1}:

$$z_k^{(2)} = z_{k+1}^{(1)} \quad , \tag{2.12}$$

and by means of Equations (2.11a) and (2.11b) we arrive at a recursion formula that, given the foreground $(k = 1)$ distance z_1, allows us to compute the optimal set of focal distances:

$$z_{k+1} = \left(\frac{1}{z_k} - \frac{c}{Rd} \right)^{-1} . \tag{2.13}$$

For a given 3D scene, the foreground distance z_1 can be determined by adjusting the current through the ETL until the foreground of the scene is imaged in-focus at the camera sensor. The value of the current, by means of linear interpolation on the calibration curve in Figure 2.6, gives z_1 (since $fl_1 = z_1$). For example, for the case of the pins set $z_1 = 66$mm and the optimal stack of $N = 22$ images covers in depth the region of interest (see *Visualization acquired pins optimal stack*).

Chapter 3

All-in-focus image reconstruction

In the previous chapter we dealt with the acquisition of images, now we will focus on how to reconstruct an all-in-focus image from the multi-focus stack. All-in-focus image reconstruction is the process of combining the in-focus information contained in each image of a set of differently focused images into a single image. This process is of interest for many different areas where the optics limits the depth-of-field, *e.g.* general photography [Allen and Triantaphillidou, 2012], scientific microscopic imaging [Sarder and Nehorai, 2006], intraoperative surgery [Feruglio et al., 2013] and machine vision for industrial inspection [Harding, 2016] to name a few, and it is known under several names: multi-focus image fusion, extended depth-of-field (EDOF) and focus stacking.

Most of multi-focus image fusion methods that can be found in the literature have been tested on mildly defocused image stacks of generally only two images (*i.e.* foreground and background). Spatial domain-based methods fuse the input images according to some focus measure criterion (spatial gradient, laplacian, etc.). This type of methods derive from edge detection, hence they strongly depend on the content of the images to fuse and work well for highly textured objects with local features, but tend to fail for smoother scenes with uniform regions [Aslantas and Pham, 2007, Mahmood and Choi, 2010, Huang and Jing, 2007, Tian et al., 2011]. In this category, two of the most recent schemes correspond to Pertuz et al. [Pertuz et al., 2013], who propose an adaptive method that is able to reduce noise while preserving image features in the fused image, and Zhou and Wang [Zhou et al., 2014], who develop a multi-scale weighted gradient-based fusion (MWGF) scheme that can handle misregistration problems and anisotropic blur.

Multi-scale decomposition approach [Pajares and De La Cruz, 2004, Guo et al., 2012, Li et al., 2011] is based on the decomposition of the source images into a multi-scale transform domain. The images are decomposed into approximation layers

containing large scale variations in intensity, and detail layers, containing small scale variations in intensity. Then the decomposition is followed by a combination of the transformed approximation and detail coefficients coming from different images with a given fusion rule to obtain new coefficients. The fused image is realized by performing the inverse transform over those new coefficients. Recently, Liu et al. [Liu et al., 2015] studied the performance of some of the most popular multi-scale transform methods, ranging from Laplacian pyramid (LP) [Li et al., 2011, Burt and Adelson, 1983] to nonsubsampled contourlet transform (NSCT) [Zhang and Guo, 2009], in combination with sparse representation (SR) [Yang and Li, 2010] and they reported that the best performance for multi-focus image fusion is obtained under a combination of NSCT and SR with decomposition level one (NSCTSR-1).

An alternative to trying to identify the in-focus features within the input images of the multi-focus stack, consists of characterizing the optical imaging system responsible for the degradation of the acquired images. This means either measuring or modeling its impulse response (see section 3.1) in an attempt to revert the produced degradation.

3.1   Imaging a 2D scene: Point-spread function

In many practical situations, when imaging plane objects (2D scenes) that lie orthogonal to the optical axis, it is possible to consider the optical system as a linear-shift invariant system. Shift invariance means that the output of an optical system is the same at all spatial points. However, this is fundamentally not true for aberrated optical systems, which means that the linear shift-invariant systems theory needs to be used with care. If a system is linear then it can be completely characterized by its response to unit impulses; this means that the image of a continuous object can be described by specifying the images of point sources located throughout it. For shift-invariant or space-invariant systems the image of a point source object changes only in location, not in functional form, as the point source explores the object.

Let us start by considering a plane object orthogonal to the optical axis, whose intensity distribution is given by $f(x_o, y_o)$ in the object plane. An ideal optical system would have a perfect image quality and then the ideal image of the object, whose intensity distribution will be denoted by $i(x_i, y_i)$, would be a magnified replica of the object intensity distribution, with all the details preserved:

$$i(x_i, y_i) = \frac{1}{M^2} f\left(\frac{x_i}{M}, \frac{y_i}{M}\right), \tag{3.1}$$

where the coordinates (x_i, y_i) are measured in the image plane (where the image

is formed according to the thin lens law in Equation (2.1)) and magnification M is given by $M = -\frac{z_i}{z_o}$, where z_o is the distance of the object plane from the lens and z_i is the distance of the image plane from the lens. Aside from the factor $\frac{1}{M^2}$, Equation (3.1) implies the replication at (x_i, y_i) of the intensity distribution value at object plane point $\left(x_o = \frac{x_i}{M}, y_o = \frac{y_i}{M}\right)$. The factor $\frac{1}{M^2}$ assures conservation of radiant energy per unit time between object and image planes for negligible absorption in the optical system.

For conceptual discussions, it is typically assumed that the imaging system has unit magnification [Iizuka, 2013], so that we can directly take $f(x_i, y_i)$ as the object irradiance distribution, albeit as a function of image-plane coordinates. Then, eq. (3.1) for ideal imaging can be put in the following form:

$$i(x_i, y_i) = f(x_i, y_i). \tag{3.2}$$

Now, a continuous intensity distribution in the object plane can be thought as a collection of point sources (delta functions) at specific locations, each with a strength proportional to the object intensity at that particular location. Any given point source has a weighting factor that can be expressed using the sifting property of the delta function:

$$f(x_i, y_i) = \iint\limits_{-\infty}^{+\infty} \delta(x_i - x_o, y_i - y_o) f(x_o, y_o) dx_o dy_o. \tag{3.3}$$

Then, Equation (3.2) can be expressed as:

$$i(x_i, y_i) = \iint\limits_{-\infty}^{+\infty} \delta(x_i - x_o, y_i - y_o) f(x_o, y_o) dx_o dy_o. \tag{3.4}$$

The previous result is valid for an ideal imaging system, where a point source in the object plane, mathematically represented by a delta function, will be imaged as another delta function in the imaging plane (*i.e.* the response of the imaging system to an impulse is another impulse). But for real systems the system quality is degraded and to describe the response of the system to an impulse, we need to consider the point spread function (PSF) of the system (also known as the impulse response of the system): $h(x_i, y_i)$. If we assume that the optical system is lossless

(*i.e.* no light energy is absorbed) then

$$\iint\limits_{-\infty}^{+\infty} h(x,y)dxdy = 1,\tag{3.5}$$

because the light energy incident on the lens was taken to be unit. Hence, for a real imaging system Equation (3.4) should be rewritten in terms of the normalized impulse response:

$$i(x_i, y_i) = \iint\limits_{-\infty}^{+\infty} h(x_i - x_o, y_i - y_o)f(x_o, y_o)dx_o dy_o,\tag{3.6}$$

The previous expression is recognized as a two-dimensional convolution of the object intensity distribution with the impulse response of the system (see Figure 3.1). In short hand notation,

$$i(x_i, y_i) = f(x_i, y_i) * h(x_i, y_i) = h(x_i, y_i) * f(x_i, y_i),\tag{3.7}$$

where the $*$ symbol means 2D-convolution.

From now on, to simplify the notation we will consider x and y instead of x_i and y_i, so Equation (3.7) reads:

$$i(x, y) = h(x, y) * f(x, y).\tag{3.8}$$

If we know the PSF, we can calculate how a 2D object will be imaged by the optical system. The convolution with the PSF blurs the image and reduces its sharpness. Convolution with the PSF in the space domain is a quite complex operation while in Fourier space (frequency domain), however, it is much simpler to perform. The Fourier transform (also Fourier spectrum or frequency spectrum) of a function $g(x, y)$ of two independent variables x and y (in space domain) will be represented by

$$G(u, v) = \mathcal{F}\{g(x, y)\} = \iint\limits_{-\infty}^{+\infty} g(x, y)exp[-j2\pi(ux + vy)]dxdy\tag{3.9}$$

where (u, v) are spatial frequencies. The convolution theorem [Goodman, 1996] states that the Fourier transform of the convolution of two functions in the space domain results in the product of their individual transforms in frequency domain:
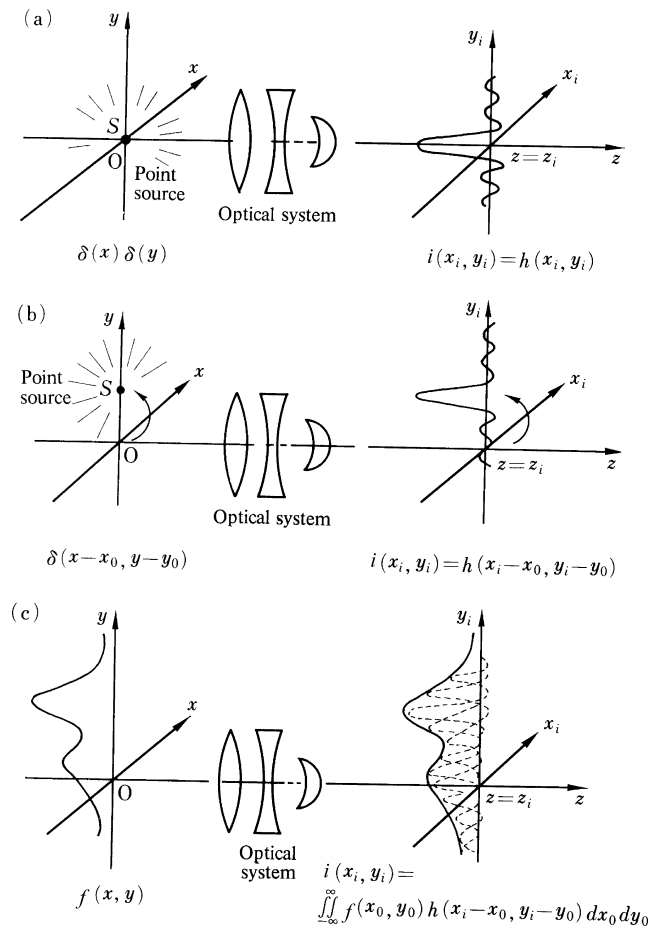
Figure 3.1: Impulse response $h(x, y)$. a) Point source at origin in object space, b) point source at (x_o, y_o), c) distributed source. Adapted from [Iizuka, 2013]

37

Convolution theorem. If $\mathcal{F}\{f(x,y)\} = F(u,v)$ and $\mathcal{F}\{g(x,y)\} = G(u,v)$, then

$$\mathcal{F}\{\iint\limits_{-\infty}^{+\infty} f(\xi,\eta)g(x-\xi, y-\eta)d\xi d\eta\} = F(u,v)G(u,v) \tag{3.10}$$

then Fourier transforming both sides of Equation (3.8) yields:

$$I(u,v) = H(u,v)F(u,v), \tag{3.11}$$

where I, H and F are the Fourier transform of i, h and f, respectively (*i.e.* $I = \mathcal{F}\{i\}$, $H = \mathcal{F}\{h\}$ and $F = \mathcal{F}\{f\}$). In this way the convolution of two functions in the space domain is equivalent to the simpler operation of multiplying their individual transforms and inverse transforming. The Fourier transform H of the impulse response h is called the *optical transfer function* OTF of the system, which among other properties verifies [Goodman, 1996]:

  (i) $H(0,0) = 1$

 (ii) $H(-u,-v) = H^*(u,v)$

(iii) $|H(u,v)| \leq |H(0,0)|$

3.2 Imaging a 3D scene: depth-invariant and depth-variant PSF

Linearity and space invariance in PSF is a desired condition for applying optical-transfer theory and the simplifications introduced by convolution operations. Image formation of 3D scenes has been modeled mathematically as the 3D convolution of a 3D PSF of the system with the 3D intensity distribution of the underlying object $f(x,y,z)$. Thus the retrieval of the object distribution is accomplished by means of 3D deconvolution techniques (depth-invariant image restoration) assuming a known 3D PSF for the optical system. In particular, this approach has been extensively used for image fusion in 3D microscopy.

However, most 3D Image deconvolution algorithms assume a 3D linear space-invariant point spread function (3D PSF) (*i.e.* shift-invariant in (x,y) plus depth-invariant in z) which is a simplification of the PSF modelling [Sarder and Nehorai, 2006, Liu and Hua, 2011]. It results from an approximation of small defocus or from imaging a relatively thin sample (*i.e.* thinner than the DOF of the optical system) where the depth variance of the PSF can be ignored, but for relatively thick samples

or strongly defocused images, this condition does not hold [Preza and Conchello, 2004, Kim and Naemura, 2015].

In alternative approaches for thick specimens, the object space is stratified into non-overlapping strata following the assumption that the 3D PSF is invariant throughout each layer or stratum. The number of strata controls the tradeoff between computational complexity of the restoration algorithm and accuracy of the restoration. The strata model approach considers, instead of an unique 3D PSF, as many 3D PSFs as strata for the object space. Each stratum is retrieved as the linear interpolation of 3D convolutions of the strata with the 3D PSFs followed by a weighted average [Preza and Conchello, 2004]. As this approach strongly depends on the number of layers considered, recently variations considering Principal Component Analysis (PCA) have been introduced to reduce the number of 3D PSFs required along with the computation load [Ghosh et al., 2014, Patwary and Preza, 2015]. However, this may lead to blur discontinuity artifacts in the resulting image.

Other methods consider an overlapping domain decomposition strategy as well as the introduction of appropriate functions to smooth transitions [Hadj and Blanc-Féraud, 2012]. Hence, depth-invariant PSF is a good approximation for thin samples but not for thick specimens where a depth-variant PSF modeling is becoming an important and active area of research.

3.3 Proposed all-in-focus image reconstruction in Fourier Domain

In the literature there has been little emphasis on all-in-focus reconstruction from strongly defocused images, although this is a especially important case in scientific microscopy. We propose a new physically based method with a depth-variant point spread function (PSF) to accomplish all-in-focus reconstruction (image fusion) from a multi-focus image sequence in order to extend the depth-of-field of a 3D scene [Alonso et al., 2015]. The proposed method works well under strong defocusing conditions for color image stacks of arbitrary length. Experimental results are provided to demonstrate that our method outperforms state-of-the-art image fusion algorithms for severe defocus on both synthetic as well as real images.

When a camera forms an image of a three-dimensional scene, it necessarily discards certain information about that scene. This loss of information is a direct result of the perspective projection that reduces the dimensionality from three to two [Kenneth, 1996]. Then, it is natural to extent the two-dimensional model for an image of a plane object parallel to the lens given by Equation (3.8) to a set of two-dimensional differently focused images of a three-dimensional object or scene as

a piecewise-planar approximation to the three-dimensional imaging space.

Consider for example a set of images of a scene with different in-focus planes acquired varying the focal length of the system (for example by means of the ETL-camera system described in Chapter 2). Each image will contain both in-focus components of the scene as well as out-of-focus contributions. Let i_k be the intensity distribution of the k-th image of a stack of N images. (For color images in RGB space $i_k = \left(i_k^R, i_k^G, i_k^B \right), k = 1, ..., N$). The i_k image taken with the system focused at axial distance $z = z_k$ can be mathematically described, neglecting noise and chromatic aberrations (that is, same response regardless of light wavelength), by the following equation:

$$i_k(x, y) = f_k(x, y) + \sum_{k' \neq k} h_{kk'}(x, y) * f_{k'}(x, y), \tag{3.12}$$

where f_k is the in-focus region of i_k. The part of the scene that looks defocused in i_k comes from the 2D convolution between $f_{k'}$ (in-focus part of $i_{k'}$ for each color channel) and the 2D intensity PSF $h_{kk'}(x, y)$ associated with the depths z_k and $z_{k'}$. Figure 3.2 shows this for $N = 2$ images. The example in Fig. 3.3 shows the optical system focused at z_k and the contributions to the image i_k from a focused red push pin and a defocused green pin at $z_{k'}$.
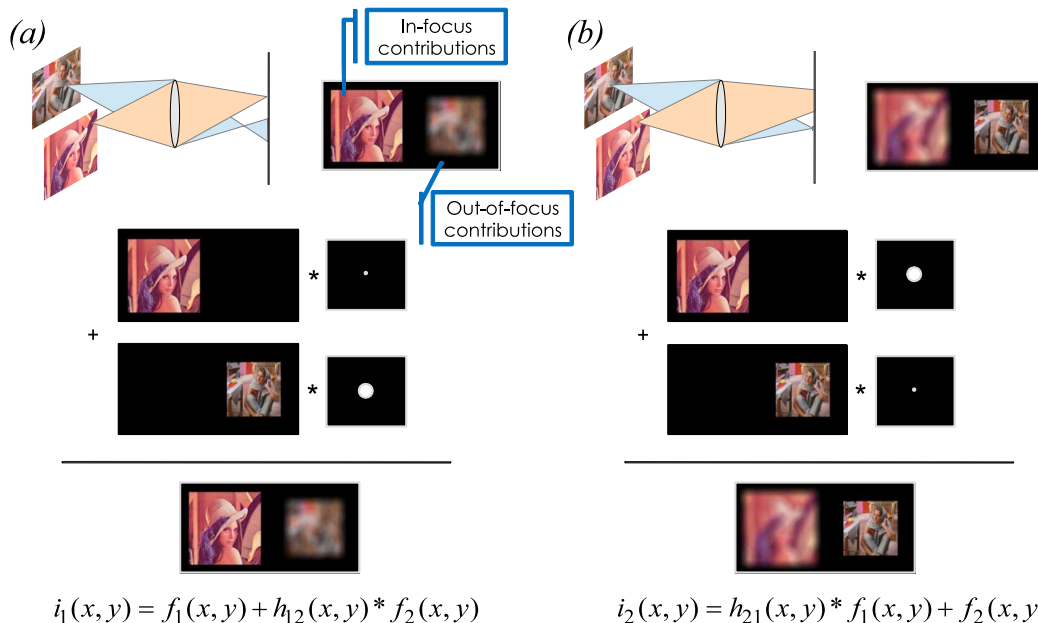
Figure 3.2: In-focus and out-of-focus contribution to image formation for $N = 2$ with the optical system focusing at (a) foreground and (b) background

When severe defocus is present, the PSF is determined primarily by geometrical-optics effects, and diffraction plays a negligible role in determining its shape [Good-
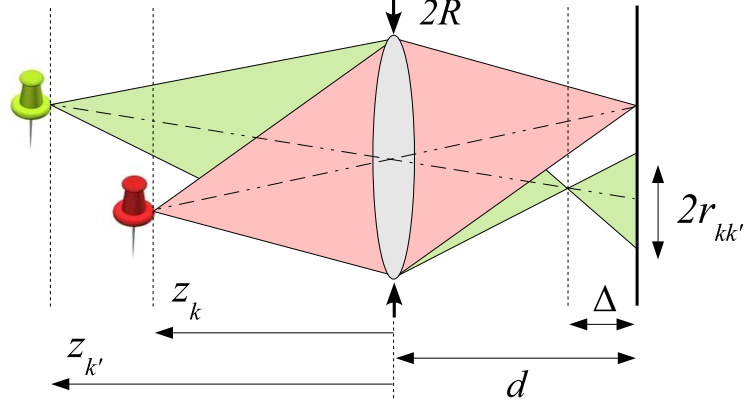
Figure 3.3: Optical system focusing at z_k (the red push pin is in focus, and the green one at $z_{k'}$ is out-of-focus).

man, 1996, Mahajan and Díaz, 2016]. The (normalized intensity) PSF in Equation (3.5) is then given by the geometrical projection of the exit pupil of the system onto the imaging plane and therefore it should be uniformly bright over a circle and zero elsewhere. If the radius of this blur circle is $r_{kk'}$, we obtain the (normalized) PSF to be a circular function [Subbarao and Surya, 1994, Aslantas and Pham, 2007]:

$$h_{kk'}(x,y) = \frac{1}{\pi r_{kk'}^2} circ \left(\frac{\sqrt{x^2+y^2}}{r_{kk'}} \right), \tag{3.13}$$

where the circle function is given by:

$$circ \left(\frac{\sqrt{x^2+y^2}}{r_{kk'}} \right) = \begin{cases} 1, & \sqrt{x^2+y^2} \leq r_{kk'} \\ \\ 0, & otherwise \end{cases} \tag{3.14}$$

The OTF corresponding to the PSF in Equation (3.13) is given by its Fourier Transform:

$$H_{kk'}(u,v) = \frac{2J_1(2\pi r_{kk'}\sqrt{u^2+v^2})}{2\pi r_{kk'}\sqrt{u^2+v^2}}, \tag{3.15}$$

where J_1 is the first order Bessel function. For the proposed system the radius of the blur circle $r_{kk'}$ can be calculated as follows: Considering similar triangles in Figure 3.3:

$$\frac{r_{kk'}}{\Delta} = \frac{R}{d-\Delta}, \tag{3.16}$$

that is:

$$r_{kk'} = R \left| \frac{\Delta}{d-\Delta} \right|, \tag{3.17}$$

41

where we consider the possible case $\Delta < 0$ (not depicted in Figure 3.3) by taking the absolute value of the right hand side. From the Gauss lens law we have

$$\frac{1}{z_{k'}} + \frac{1}{d - \Delta} = \frac{1}{fl} \tag{3.18a}$$

$$\frac{1}{z_k} + \frac{1}{d} = \frac{1}{fl} \tag{3.18b}$$

where fl is the effective focal length for the system focusing at z_k. Equating the left hand sides of Equations (3.18a) and (3.18b) we obtain

$$\frac{1}{z_{k'}} + \frac{1}{d - \Delta} = \frac{1}{z_k} + \frac{1}{d} \quad , \tag{3.19}$$

which is equivalent to

$$\frac{1}{z_k} - \frac{1}{z_{k'}} = \frac{1}{d}\left(\frac{\Delta}{d - \Delta}\right) \quad , \tag{3.20}$$

and substituting in Equation (3.17) we finally obtain

$$r_{kk'} = Rd\left|\frac{1}{z_k} - \frac{1}{z_{k'}}\right|. \tag{3.21}$$

It is clear from Equation (3.21) that the PSF under strong defocus cannot be considered as space invariant on the axial coordinate since $r_{kk'}$ does not depend solely on the difference between z_k and $z_{k'}$[1]. Often, in image processing algorithms, instead of resembling the geometrical shape of the aperture of the lens, the PSF of a system under white light is modeled as a 2D Gaussian function (basically to simplify the calculus). However, observed PSF can be quite different from 2D Gaussian, and therefore the Gaussian is not a satisfactory model [Subbarao, 1990]. In photography, for example, this geometric effect is known as *bokeh*. The key to using *bokeh* in a shot is to use a wide aperture on a close focused subject in a way that bright out-of-focus elements in the background resemble the shape of the diaphragm (the aperture) of the lens (see Figure 3.4). A lens with more circular shaped blades will have rounder, softer orbs of out-of-focus highlights, whereas a lens with an aperture that is more hexagonal in shape will reflect that shape in the highlights [2].

In order to computationally implement the algorithm, the radius of defocus $r_{kk'}$

---

[1] Only for small defocusing values or under thin object approximation, Δ is linear with $z_{k'} - z_k$ and the PSF can be considered invariant on the axial coordinate (*i.e.* depth-invariant) [Kirshner et al., 2013].

[2] http://www.nikonusa.com/en/learn-and-explore/article/h0ndz86v/bokeh-for-beginners.html

Figure 3.4: Image taken with a camera focusing at the glasses in the foreground. The out-of-focus highlights in the background resemble the geometrical shape of the aperture (in this case the diaphragm opening in the camera). In photography this in known as *bokeh*. Photo credit: Bryan Leung http://www.nikonusa.com/en/learn-and-explore/article/h0ndz86v/bokeh-for-beginners.html

in Equation (3.21) should be expressed in terms of pixel pitch p of the camera:

$$\frac{r_{kk'}}{p} = R_0 \left| \frac{1}{z_k} - \frac{1}{z_{k'}} \right|, \tag{3.22}$$

where $R_0 = \frac{Rd}{p}$, R is the radius of the imaging lens, and d is the distance from this lens to the imaging plane (see Figure 3.3). R_0 is the only parameter required to get the fused image from the stack.

The fused image $s(x, y)$ will correspond to the sum of the in-focus components of the different images:

$$s(x, y) = \sum_{k=1}^{N} f_k(x, y). \tag{3.23}$$

In order to obtain $s(x, y)$, we consider the Fourier transform ($\mathcal{F}$) of Equation (3.12):

$$I_k(u, v) = F_k(u, v) + \sum_{k' \neq k} H_{kk'}(u, v) F_{k'}(u, v), \tag{3.24}$$

where (u, v) are spatial frequencies, $H_{kk'}$ is the optical transfer function (OTF) (*i.e.* the Fourier transform of $h_{kk'}$: $H_{kk'} = \mathcal{F}\{h_{kk'}\}$), while $I_k = \mathcal{F}\{i_k\}$ and $F_{k'} = \mathcal{F}\{f_{k'}\}$ (the Fourier transforms are implemented for each color channel).

For the particular case $N = 2$, the system for $k = 1, 2$ in Equation (3.24) takes the form:

$$I_1(u, v) = F_1(u, v) + H_{12}(u, v)F_2(u, v) \qquad (3.25a)$$

$$I_2(u, v) = H_{12}(u, v)F_1(u, v) + F_2(u, v). \qquad (3.25b)$$

By adding the previous equations and taking into account Equation (3.23), we get

$$S(u, v) = \mathcal{F}\{s(x, y)\} = F_1(u, v) + F_2(u, v) = \frac{I_1(u, v) + I_2(u, v)}{1 + H_{12}(u, v)}, \qquad (3.26)$$

and the fused image s can be obtained from

$$s(x, y) = \mathcal{F}^{-1}\{S(u, v)\} = \mathcal{F}^{-1}\left\{\frac{I_1 + I_2}{1 + H_{12}}\right\}. \qquad (3.27)$$

Other approaches to a linear treatment of the case of two depth layers (foreground and background), $N = 2$, can be found in [Schechner et al., 2000, Kubota and Aizawa, 2005].

For arbitrary N, the set of coupled equations in Equation (3.24) can be arranged in vector form as

$$\vec{I}(u, v) = H(u, v)\vec{F}(u, v), \qquad (3.28)$$

where N-element column vectors $\vec{I}$, $\vec{F}$ and $N \times N$ symmetric matrix H are, respectively, given by

$$\vec{I}(u, v) = \begin{pmatrix} I_1(u, v) \\ I_2(u, v) \\ \vdots \\ I_N(u, v) \end{pmatrix}, \quad \vec{F}(u, v) = \begin{pmatrix} F_1(u, v) \\ F_2(u, v) \\ \vdots \\ F_N(u, v) \end{pmatrix},$$

$$\tag{3.29}$$

$$H(u, v) = \begin{pmatrix} 1 & H_{12}(u, v) & \dots & H_{1N}(u, v) \\ & 1 & & \vdots \\ & & \ddots & H_{N-1N}(u, v) \\ & & & 1 \end{pmatrix},$$

where the elements below the diagonal in symmetric matrix H have been omitted since they repeat the elements above the diagonal. Let us now consider the Fourier

transform of Equation (3.23), then

$$S(u,v) = \mathcal{F}\{s(x,y)\} = \sum_{k=1}^{N} F_k(u,v) = \vec{e} \cdot \vec{F}(u,v), \tag{3.30}$$

where $\vec{e}$ is the vector of length N whose elements are ones and $\cdot$ indicates inner product. Notice that in Equation (3.28) we know $\vec{I}(u,v)$ since it comes from the Fourier Transform of the acquired images and we have constructed the matrix $H(u,v)$ by modeling the 2D PSFs. Now, $H(u,v)$ is not invertible for every pair of frequencies (u,v) (we will see below that this is the case when $(u,v) = (0,0)$), then the retrieval of $\vec{F}(u,v)$ through inversion of the system given by Equation (3.28) is not always possible (*i.e.* it's an *ill-posed* problem). Then we need to compute $S(u,v)$ knowing $\vec{I}(u,v)$ and $H(u,v)$ and not knowing $\vec{F}(u,v)$.

From Equations (3.28) and (3.30), we propose to retrieve $S(u,v)$ as

$$S(u,v) = \vec{e} \cdot \left(H^\dagger(u,v)\vec{I}(u,v) \right), \tag{3.31}$$

where $H^\dagger$ is the Moore-Penrose pseudo-inverse [Ben-Israel and Greville, 2003], which reduces to H^{-1} when H is invertible. $H^\dagger$ provides the set of vectors that minimize the Euclidean norm $\left\| H(u,v)\vec{F}(u,v) - \vec{I}(u,v) \right\|$ in the least squares sense and $H^\dagger \vec{I}$ is the minimum norm vector within this minimal set (see Appendix A).

In the particular case when $(u,v) = (0,0)$, since $H_{kk'}$ are OTF-functions whose form is given by Equation (3.15) and according to Item (i) from Section 3.1: $H_{kk'}(0,0) = 1$. Thus, $H(0,0)$ is an all-ones matrix, which is non invertible. From Equation (3.30), it turns out that $S(0,0) = I_k(0,0)$ for any k (this result can also be obtained straightforwardly from Equation (3.24)). In practice, it could be better to take the average in order to compensate small differences between them, *i.e.* we calculate the DC component of S from

$$S(0,0) = \frac{1}{N}\vec{e} \cdot \vec{I}(0,0). \tag{3.32}$$

Through computation of Equations (3.31) and (3.32) (using MATLAB), we obtain S, and then, by taking its inverse Fourier transform, the desired all-in-focus image $s(x,y)$ is retrieved:

$$s(x,y) = \mathcal{F}^{-1}\{S(u,v)\}. \tag{3.33}$$

Since the DC component $I_k(0,0)$ is the radiant energy corresponding to image i_k

45

(consider $(u, v) = (0, 0)$ in Equation (3.9) applied to $i_k(x, y)$),

$$I_k(0,0) = \iint\limits_{-\infty}^{+\infty} i_k(x,y)dxdy,$$ (3.34)

to ensure the conservation of radiant energy between images of the stack, the set of multi-focus images will be normalized by the mean DC component:

$$\vec{I}(0,0) = \frac{1}{N}\left(\sum_k I_k(0,0)\right)\vec{e}.$$ (3.35)

3.4 Comparison with state-of-the-art methods

It is desirable to generate synthetic data to test and quantify the performance of a given model. One can compute the digital image that simulates the capture of an object as seen through an optical system and also compare the all-in-focus reconstruction from different methods against a known *ground truth*[1]. To measure the reliability of the methods we tested them on synthetic defocused images and computed the *Mean Squared Error* (MSE) between the reconstructions and the *ground truth*. The MSE is a statistical measure that compares on a pixel-by-pixel basis the luminance patterns[2] from the *ground truth* image with the all-in-focus reconstructions. It is calculated through the following equation:

$$MSE(i_{aif}, i_{gt}) = \frac{1}{N_{row}N_{col}}\left(\sum_x \sum_y [\ell_{aif}(x,y) - \ell_{gt}(x,y)]^2\right).$$ (3.36)

where ℓ_{aif} and ℓ_{gt} represent the all-in-focus reconstructed and *ground truth* image's luminances for the pixel located at row y and column x, and N_{row} and N_{col} represent the number of rows and columns of the images.

The performance of the proposed method was tested on both synthetic as well as real data experiments against Pertuz et al. [Pertuz et al., 2013], *MWGF* [Zhou et al., 2014], LP (level 4) [Li et al., 2011], NSCT (level 4), and NSCT-SR-1 methods [Yang and Li, 2010]. The following considerations were taken into account for the comparison: MWGF and NSCT-SR-1 have been reported to work with stacks of two source images; Pertuz et al. is intended for large stacks and requires at least four

---

[1] *Ground truth*: reference image against which quantitative comparisons can be made.

[2] The luminance signal is the brightness of an image as displayed by a black-and-white television receiver. It is a weighted sum of the R, G, and B components in proportion to the human eye's sensitivity to them [Russ, 2016]. The *rgb2gray* MATLAB's function converts RGB images to grayscale by performing that weighted sum: ℓ=0.299 R + 0.587 G + 0.114 B [Gonzalez et al., 2009].
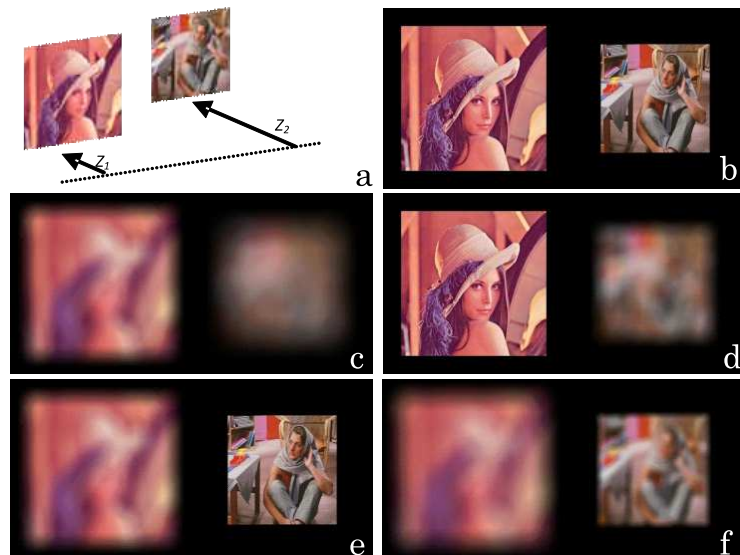
Figure 3.5: Synthetic defocused images generated from Lena and Barbara at different distances. (a) Synthetic setup, (b) reference image (*ground truth*). For $R_0 = 160z_1$, system focused at (c) $z_1 - \delta$, (d) z_1, (e) z_2, and (f) $z_2 + \delta$, with $\delta = 0.2z_1$.

source images; LP, NSCT and the proposed method can handle stacks of arbitrary length. Figure 3.5(a) shows a synthetic setup with *Lena* image at distance z_1 and *Barbara* image at distance $z_2 = \frac{4}{3}z_1$ of the imaging lens. Unlike real multi-focus stacks, we can construct the *ground truth* (Figure 3.5(b)) where both Lena and Barbara are in focus and serves as a reference to test the quality of the fused image. The defocus was implemented considering mild to strong defocus by taking $R_0 = 20z_1$, $40z_1$, $80z_1$, and $160z_1$ in Equation (3.22). For each R_0 value, we constructed a synthetic stack of $N = 4$ images, each one corresponding to the system focused at $z_1 - \delta$, z_1, z_2, and $z_2 + \delta$, respectively ($\delta = 0.2z_1$). Figures 3.5(c)-(f) show the case $R_0 = 160z_1$.

When a stack of only $N = 2$ images was necessary, we used the subset of Figures 3.5(d) and 3.5(e): Figure 3.5(d) simulates the in-focus capture of Lena placed at z_1 and the out-of-focus Barbara at z_2, while Figure 3.5(e) simulates the in-focus capture of Barbara with Lena defocused. From Equation (3.22) and using the relation between z_2 and z_1, the out-of-focus parts of each of the 2 images come from blurring with disks of radii 5, 10, 20, and 40 pixels for $R_0 = 20z_1$, $40z_1$, $80z_1$, and $160z_1$, respectively.

Figure 3.6 shows the fused images obtained by the different methods, (a_1)-(a_5) for $N = 2$ and (b_1)-(b_4) for $N = 4$. Mean square error (MSE) values (Equation (3.36))

Table 3.1: MSE for Different Degrees of Defocus for Stack Lengths of $N = 2$ and $N = 4$ Images

R_0	N	Pertuz et al.	MWGF	LP	NSCT	NSCT-SR-1	Proposed method
$20z_1$	2	-	**0.1338**	0.3942	0.7011	0.5834	0.6792
	4	**0.3083**	-	4.5960	1.6027	-	3.6250
$40z_1$	2	-	1.6220	0.5813	3.5009	5.6416	**0.5194**
	4	5.5684	-	3.8791	11.6594	-	**2.6968**
$80z_1$	2	-	3.8487	2.7873	29.4780	37.0632	**0.4438**
	4	6.8451	-	5.9722	78.5429	-	**2.5025**
$160z_1$	2	-	13.3223	31.6798	115.7629	113.8175	**0.4113**
	4	6.5985	-	76.6185	266.5993	-	**3.4948**

for the fused images obtained with the different methods can be found in Table 3.1 (comparison made between respective luminances). The values labeled in bold indicate the best performance over all the methods for the corresponding R_0 and N. It can be seen that the proposed method outperforms the other schemes as defocus increases.

To test the robustness of our method, we considered a deviation from the R_0 value used to create the defocus blur. For example, in the case $N = 2$ and defocusing due to $R_0 = 160z_1$, by taking $R_0 = 128z_1$ and $R_0 = 192z_1$ for performing the fusion (*i.e.* 20% deviation from the parameter value used to generate the defocused images), the proposed scheme gives $MSE = 12.9989$ and $MSE = 10.7235$, respectively, both still below the second best method.

We have also tested the performance of our method for real images. Figure 3.7 (which repeats Figure 2.10) shows $N = 4$ source images corresponding to four colored pins placed at different distances of the imaging plane. The images were acquired using the optical system described in Chapter 2, for which $2R = 11.5$mm, $d \approx 16$mm, and $p = 5.65\mu$m, so that $R_0 \approx 16300$mm.

Figure 3.8 shows the results of the fusion images by Pertuz et al., LP, NSCT and the proposed method. Although in real images, we lack of a ground-truth for comparison purposes, it can be qualitatively observed that the outcome of the fusion process by the proposed algorithm is very good. From the detail of the results of the different fusion methods [Figs. 3.8(e)-3.8(h)], it can be seen that the proposed method performs better than the others, where overshoot, halo artifacts, or contrast degradation are observed.

Most of the methods to extend the depth-of-field are implemented for only two images (foreground and background) and only a few can deal with more images. This is the reason that we have considered two or four images in the previous comparisons.
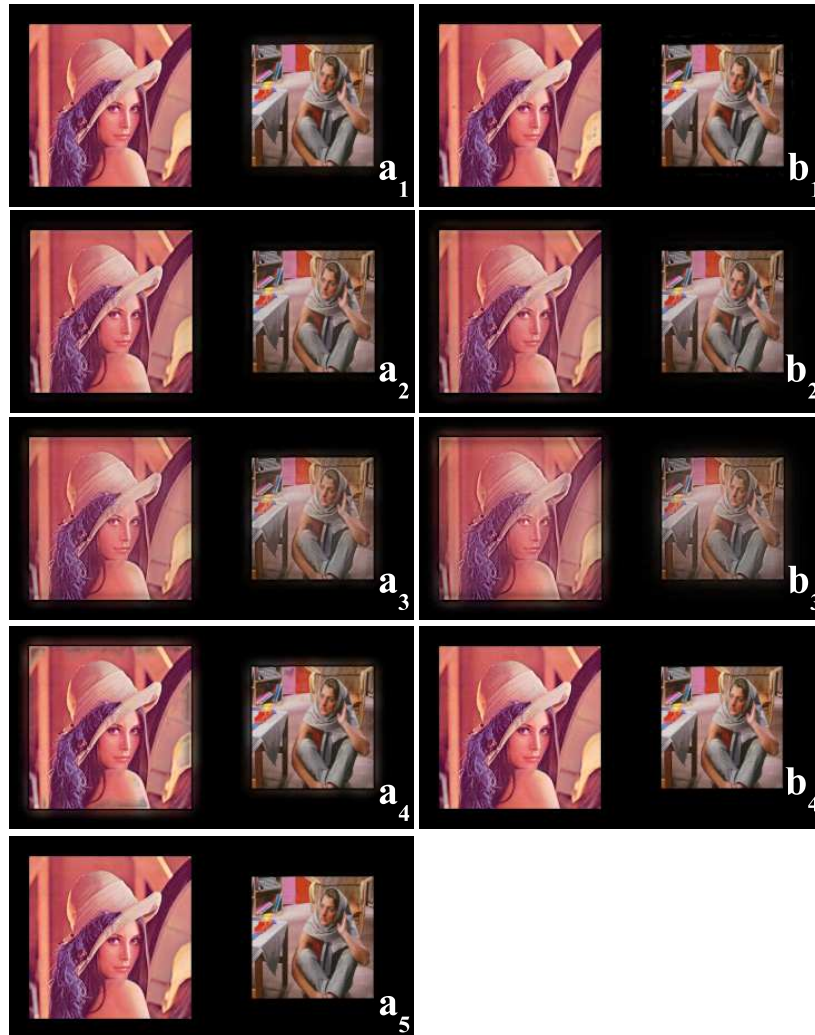
Figure 3.6: Results for $R_0 = 160z_1$ from different methods.(a) two-image stack; (b) four-image stack. From a_1 to a_5 MWGF, LP, NSCT, NSCT-SR-1, and proposed method, respectively. From b_1 to b_4 Pertuz et al., LP, NSCT, and proposed method, respectively.
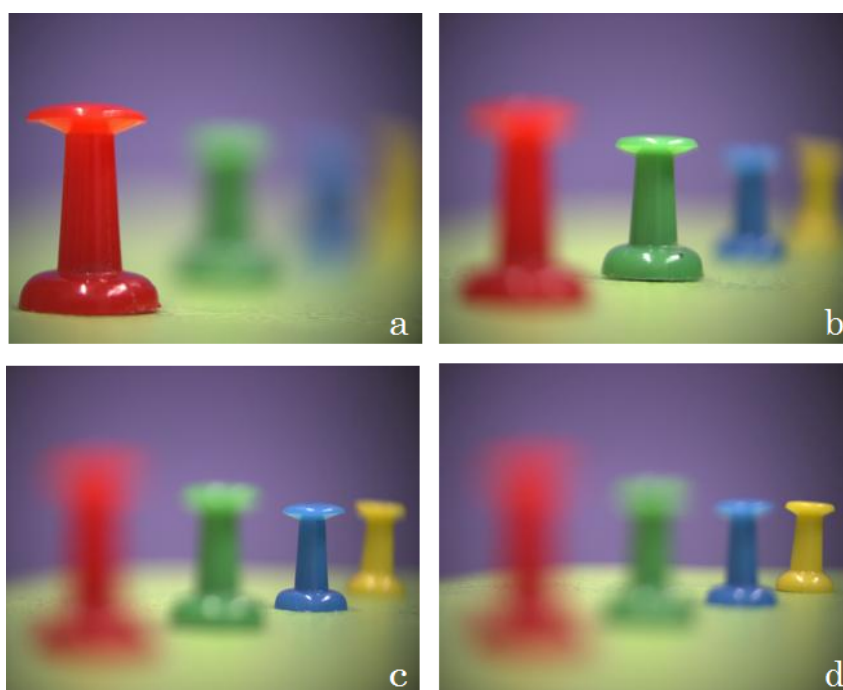
Figure 3.7: System focused at (a) red $z_k = 69$mm, (b) green $z_k = 99$mm, (c) blue $z_k = 129$mm, and (d) yellow $z_k = 159$mm push pins.
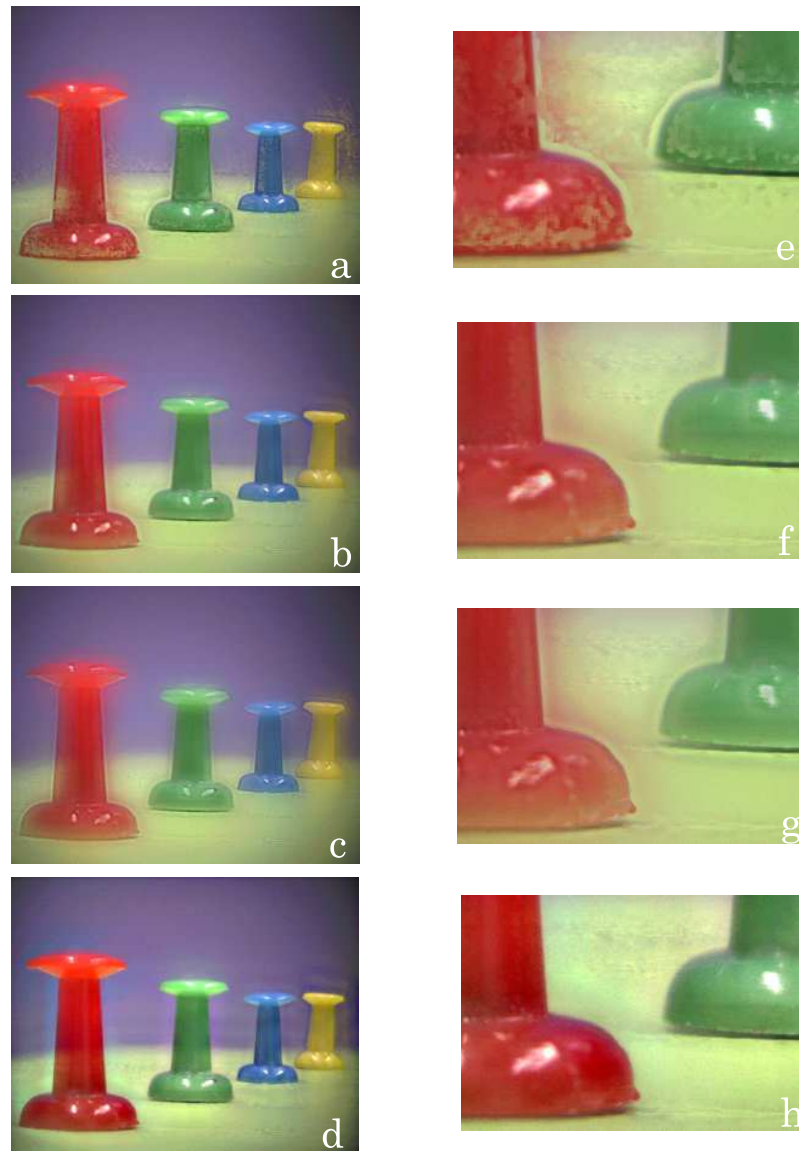
Figure 3.8: Image fusion results from source images in fig. 3.7 under (a) Pertuz et al., (b) LP, (c) NSCT, and (d) proposed method. (e)-(h) Detail of the fusion results (a)-(d), respectively.

However, in Figure 3.9, we present the result of the proposed method for four images in the pins scene (above) for the corresponding optimal stack (see Section 2.4 and *Visualization acquired pins optimal stack*) for the same scene (below). The outcome of the all-in-focus image reconstruction for the optimized acquisition, is a fused image that seems to be captured with an optical system with infinite DOF.
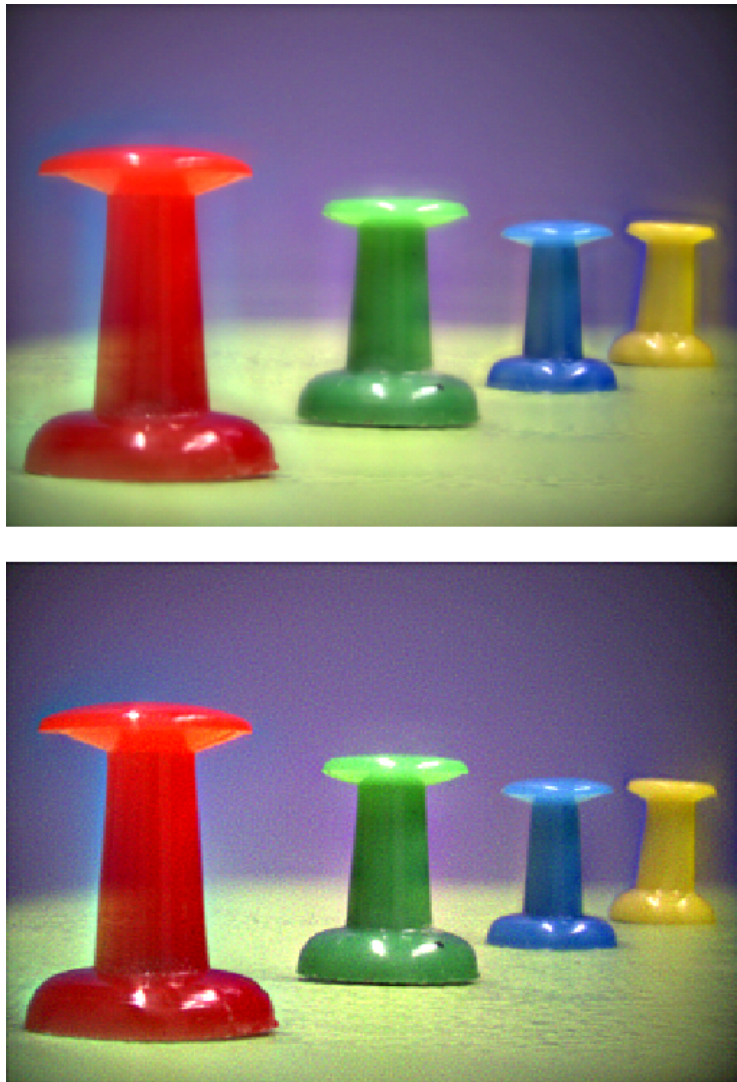
Figure 3.9: All-in-focus image fusion for the push pins scene by the proposed method implemented for four image sequence (top) and from the optimized sequence: twenty two images (bottom)

The simplicity of the method lies in the use of linear system theory and in the independence respect to the content of the input images unlike other methods which use focus measures that rely on highly textured objects or local features. Besides,

there is no need for segmentation or retrieval of the depth-map of the scene. Linear system theory can be strictly applied to an optical system under the assumption that the 3-D object can be modeled as transparent layers [Preza et al., 1992, Schechner et al., 2000, Kubota and Aizawa, 2005].

Nevertheless we processed occlusion-free images as well as images with partial occlusions. Better results are expected for multi-planar non-occluded objects or for continuous surfaces that do not present self-occlusions, while some artifacts are expected in the reconstruction of scenes where occlusions are present (usually a slight see-through effect). Dealing with occlusion is a highly complex problem and we will not treat it here. An interesting approach to the topic can be found in [Favaro and Soatto, 2003, Favaro and Soatto, 2007].

On the other hand, noise usually arises during acquisition process through a digital imaging system and results in pixel values that do not reflect the true intensities of the real scene. The combined effect of the different noise sources is often modeled by an additive noise η in Equation (3.8): $i(x,y) = h(x,y) * f(x,y) + \eta(x,y)$. However, for the considered optical system under good illumination, the signal-to-noise-ratio (SNR) for the acquired images is high enough to neglect the additive noise.

The proposed method can handle severely defocused images and can be used with any other optical system provided that it is calibrated to know the focusing distances at which the images were acquired (this information is usually saved in the metadata as the subject to lens distance), and that registration between the acquired images is done in a preprocessing step, if needed. Finally, although the segmentation of the in-focus regions, known as *focus slicing*, is not an objective of the method, it is discussed for completeness in [Alonso and Ferrari, 2015], see Appendix B.

Chapter 4

Refocusing by post-capture aperture reshaping

How pupil shape and size affect image capture is an active investigation topic in many research areas, *e.g.* coded aperture [Nugent, 1987, Levin et al., 2007], variable aperture control [Tsai and Yeh, 2010, Xu et al., 2015, Schuhladen et al., 2013], and asymmetrical apertures for focal side detection [Sellent and Favaro, 2014]. As we will see below, there is an orientation flip of the PSF which is visible for asymmetric apertures around the focusing plane which makes these pupils more sensitive to defocus [Greengard et al., 2006, Jesacher et al., 2015]. Also these kind of pupils seem to play an important role in the evolution of some animals eye's pupil shape [Held et al., 2012]. A coded aperture is a particular pattern of transparent and opaque regions designed for defocus deblurring [Zhou et al., 2011]. To study the effects of a particular coded aperture, usually the aperture of a camera is either modified by physically interchanging fixed pattern masks [Veeraraghavan et al., 2007], or it can be dynamically changed through a programmable liquid crystal array [Liang et al., 2008, Nagahara et al., 2010].

As we saw in Chapter 3 the shape of the aperture determines the shape in which the out-of-focus points are blurred in the image (*i.e.* the PSF) as well as the corresponding depth of field (DOF). This principle is also used in photography with aesthetic purposes, sometimes taking advantage of secondary effects resulting from the design of some lenses and other times by physically changing the shape of the aperture. For example, when employing catadioptric systems, the positioning of lenses and mirrors allows the entire lens to be cheaper, lighter and smaller when compared to a traditional telephoto lens with equal focal length [Ray, 2015]. However, the annular shape of the entrance pupil causes out-of-focus highlights to take on a characteristic *doughnut* or annular shape. An example of this effect is shown in

Figure 4.1: Left: the effect of the annular shaped aperture on the defocused background due to a catadioptric lens. Photo credit: Klauss Schmitt (http://forum.mflenses.com/viewtopic.php?t=29778). Right: catadioptric lens basic structure (http://photographymc.blogspot.com.uy/2012/01/camera-lens-photography-course-lesson-2.html).

Figure 4.1. There are also commercial *bokeh kits* (https://www.amazon.com/Bokeh-Masters-Kit-Edition/dp/B004PV8N1I) where filters with different bokeh shapes are cut out. In addition, a few years ago, Sony released a *bokehmorphic* lens to allow photographers to control the shape of out-of-focus highlights in the background landscapes (see Figure 4.2).

Post-capture scene refocusing (DOF control) has also become a topic of interest in computational photography [Kodama and Kubota, 2013, Jacobs et al., 2012] and microscopic imaging [Zheng et al., 2011, Jesacher et al., 2015, Orth and Crozier, 2012]. In particular, plenoptic or *Light field* cameras [Ng et al., 2005] have had a growing popularity attributed to capabilities such as producing several images refocused at different distances from a single shot of a scene. However, they still suffer from a significant limitation, *i.e.* low spatial resolution output images (one pixel per microlens) [Bishop and Favaro, 2012].

The purpose of this chapter is to show that through the proposed Fourier Domain formalism, we are also capable of post-capture aperture reshaping and refocusing a 3D scene from a multi-focus stack, *i.e.* to reconstruct images as if they had been acquired with different pupils [Alonso et al., 2016, Alonso, 2016c]. Partially extended depth-of-field and all-in-focus image reconstruction are obtained as particular cases.

Figure 4.2: Left: SLR Bokehmorphic Sony NEX camera and lens with different diaphragms shapes, the star shaped diaphragm is shown in particular. Right: the effect of the star shaped diaphragm on the defocused highlights in the background. Photo credit: Chris Gampat (http://www.thephoblographer.com/2012/04/03/review-slr-magic-bokehmorphic-sony-nex-cameras/#.V-l_88kYGSo).

4.1 Synthetically reshaped pupils

As we already saw in the previous chapter, the set of acquired images $\{i_k, k = 1, ..., N\}$ is given by (recall Equation (3.12)):

$$i_k(x, y) = \sum_{k'=1}^{N} h_{kk'}(x, y) * f_{k'}(x, y),$$ (4.1)

where (recall Equation (3.13))[1]:

$$h_{kk'}(x, y) = \frac{1}{\pi r_{kk'}^2} circ\left(\frac{\sqrt{x^2 + y^2}}{r_{kk'}}\right),$$ (4.2)

with $r_{kk'}$ being the radius of the blurring circle resembling the circular aperture of the optical system (recall Equation (3.21)):

$$r_{kk'} = Rd\left|\frac{1}{z_k} - \frac{1}{z_{k'}}\right|.$$ (4.3)

---

[1] $h_{kk'(x,y)}$ approximates a delta function $\delta(x, y)$ when $r_{kk'}$ approximates zero (which in turn is achieved for $k' = k$); then, the term $k' = k$ in Equation (4.1) is $h_{kk}(x, y) * f_k(x, y) = f_k(x, y)$ and we can recover Equation (3.12).

If we were to acquire a set of images through a different aperture (an aperture with a different shape) then instead of the PSFs $\{h_{kk'}\}$, we would have to consider in the image formation Equation (4.1) different PSFs $\{g_{kk'}\}$ associated with the new aperture. In this case, with the system focusing at distance z_k, the acquired image $i_k^{(g)}$ will be given by:

$$i_k^{(g)}(x,y) = \sum_{k'=1}^{N} g_{kk'}(x,y) * f_{k'}(x,y). \tag{4.4}$$

Since we already have the original set of acquired images $\{i_k\}$, we are seeking to simulate the effect of a pupil (that leads to $g_{kk'}$) placed at the same position of the original physical circular aperture. Then we attempt to reconstruct the Fourier transform of $i_k^{(g)}$ from the set of acquired images $\{i_k\}$ and the focusing distances $\{z_k\}$ associated with them through the original PSFs $\{h_{kk'}\}$:

$$I_k^{(G)}(u,v) = \sum_{k'=1}^{N} G_{kk'}(u,v) \left(H^{\dagger}(u,v)\vec{I}(u,v) \right)_{k'}, \tag{4.5}$$

where $G_{kk'}$ is the Fourier Transform of $g_{kk'}$, $(G_{kk'} = \mathcal{F}\{g_{kk'}\})$ and $H^{\dagger}(u,v)\vec{I}(u,v)$ is the minimum norm solution to Equation (3.28). In vectorial form the previous equation can be arranged as:

$$I_k^{(G)}(u,v) = \vec{G}(u,v) \cdot \left(H^{\dagger}(u,v)\vec{I}(u,v) \right), \tag{4.6}$$

where $\vec{G} = (G_{k1}, G_{k2}, ..., G_{kN})^{\mathsf{T}}$, being $G_{kk'}$ the Optical Transfer Function (OTF) relating the distances z_k and $z_{k'}$ for the synthetic pupil. Then, the refocused image $i_k^{(g)}$ will be given by

$$i_k^{(g)}(x,y) = \mathcal{F}^{-1}\{I^{(G)}\} = \mathcal{F}^{-1}\left\{ \vec{G}(u,v) \cdot \left(H^{\dagger}(u,v)\vec{I}(u,v) \right) \right\}. \tag{4.7}$$

In the next we will consider different synthetic pupils, starting from the case of a pinhole.

4.1.1 Pinhole pupil

Notice that for the particular case when the 2D aperture resembles a pinhole aperture the PSFs $g_{kk'}(x,y)$ can be approximated as

$$g_{kk'}(x,y) = \delta(x,y) \quad \forall k, k'. \tag{4.8}$$

Since $\mathcal{F}\{\delta(x,y)\} = 1$ vector $\vec{G}$ will be the $N \times 1$ all ones vector $\vec{e}$, and Equation (4.6) reduces to the all-in-focus expression in Fourier domain $S(u,v)$ already obtained in Equation (3.31), regardless of the value of k[1]:

$$I^{(\delta)}(u,v) = \vec{e} \cdot \left(H^{\dagger}(u,v)\vec{I}(u,v) \right) = S(u,v). \qquad (4.9)$$

By inverse Fourier transforming the previous expression we get the all-in-focus reconstructed image (recall Equation (3.33)):

$$i^{(\delta)}(x,y) = \mathcal{F}^{-1}\{S(u,v)\} = s(x,y). \qquad (4.10)$$

Hence, the all-in-focus image is a particular case of refocusing with the aperture reshaped as a pinhole pupil. However, it is important to mention that this pinhole imaging model is clearly an idealization, because in reality when the aperture decreases that much, diffraction effects become dominant and noise may become relevant also since almost no light can pass through a real pinhole. Nevertheless, the pinhole idealization, where pinhole is small but sufficiently large so that the ray optics model is valid, is frequently used as a reasonable approximation for well-focused systems with a large depth of field [Favaro and Soatto, 2007].

4.1.2 Rotation symmetric pupils

Let us now consider pupils with rotational symmetry whose PSFs can be expressed analytically. Our first example is a circular pupil whose radius $R^{(g)}$ is smaller than the original radius R of the aperture by a factor of $\alpha > 1$: $R^{(g)} = \dfrac{R}{\alpha}$. The radius of the new blurring circle will be $r_{kk'}^{(g)} = \dfrac{r_{kk'}}{\alpha}$. Hence, the pupil of radius $R^{(g)}$ will give rise to the following 2D PSF:

$$g_{kk'}(x,y) = \frac{1}{\pi \left(\dfrac{r_{kk'}}{\alpha} \right)^2} circ \left(\frac{\sqrt{x^2 + y^2}}{\left(\dfrac{r_{kk'}}{\alpha} \right)} \right). \qquad (4.11)$$

Another example is an annular pupil, which can be implemented as the normalized substraction between two circle functions, one for the external radius, R_{ext}, and the other for the inner radius, R_{inn}. Then, the radius of the external blurring circle will be given by $r_{kk'}^{(ext)} = \dfrac{R_{ext}}{R} r_{kk'}$ while the radius of the inner blurring circle will be

---

[1] the DOF is inversely proportional to the size of the aperture and then for a pinhole the 3D scene is imaged in focus regardless of the particular focusing distance z_k

given by $r_{kk'}^{(inn)} = \dfrac{R_{inn}}{R} r_{kk'}$ and the new PSF will be given by:

$$
g_{kk'}(x,y) = \frac{1}{\pi r_{kk'}{}^2 \left(\dfrac{R_{ext}{}^2 - R_{int}{}^2}{R^2} \right)} \left\{ circ \left(\frac{\sqrt{x^2 + y^2}}{\dfrac{R_{ext}}{R} r_{kk'}} \right) - circ \left(\frac{\sqrt{x^2 + y^2}}{\dfrac{R_{inn}}{R} r_{kk'}} \right) \right\},
$$
(4.12)

4.1.3 Arbitrary pupils

In general, aside from the analytical expressions of the previous section, the PSF associated to an arbitrary center-symmetric pupil shape can be put in the following form:

$$
g_{kk'}(x,y) = \frac{p_{kk'}(x,y)}{\iint p_{kk'}(x,y)dxdy}
$$
(4.13)

where $p_{kk'}(x,y)$ corresponds to the scaled projection in the camera sensor plane of the binary function associated with the particular aperture considered.

Strictly speaking, when the in-focus imaging plane lies beyond the sensor plane, the geometric projection of the aperture is inverted with respect to the in-focus image. When the pupil is center symmetric, the original blur and its flipped version cannot be distinguished. However, for asymmetric (*i.e.* non center-symmetric) apertures this flip allows to distinguish between defocused object points before or after the focusing distance of the system.

This situation is illustrated with a hollow triangle aperture in Figure 4.3. Let us consider an object placed at $z_{k'}$ and different focusing distances z_k (as it is achieved through our system in Section 2.3.3). When the focusing distance z_k equals $z_{k'}$, the PSF reduces to a δ function and a point in the object is essentially seen as a point in the sensor (Figure 4.3 (top)). For $z_k < z_{k'}$ the PSF will correspond to a scaled projection of the lens aperture that is flipped both vertically and horizontally (Figure 4.3 (middle)), while for $z_k > z_{k'}$ the PSF will correspond to a scaled projection of the lens aperture with the same orientation (Figure 4.3 (bottom)). As the in-focus image is formed inverted through the lens (see Figure 2.1), a defocused point in the image as seen in the output (*e.g.* PC monitor) will appear as a triangle with the same orientation of the aperture (Figure 4.3 (middle, right)) or flipped (Figure 4.3 (bottom, right)).

Then, for an asymmetric aperture, the expression of the PSF in Equation (4.13)
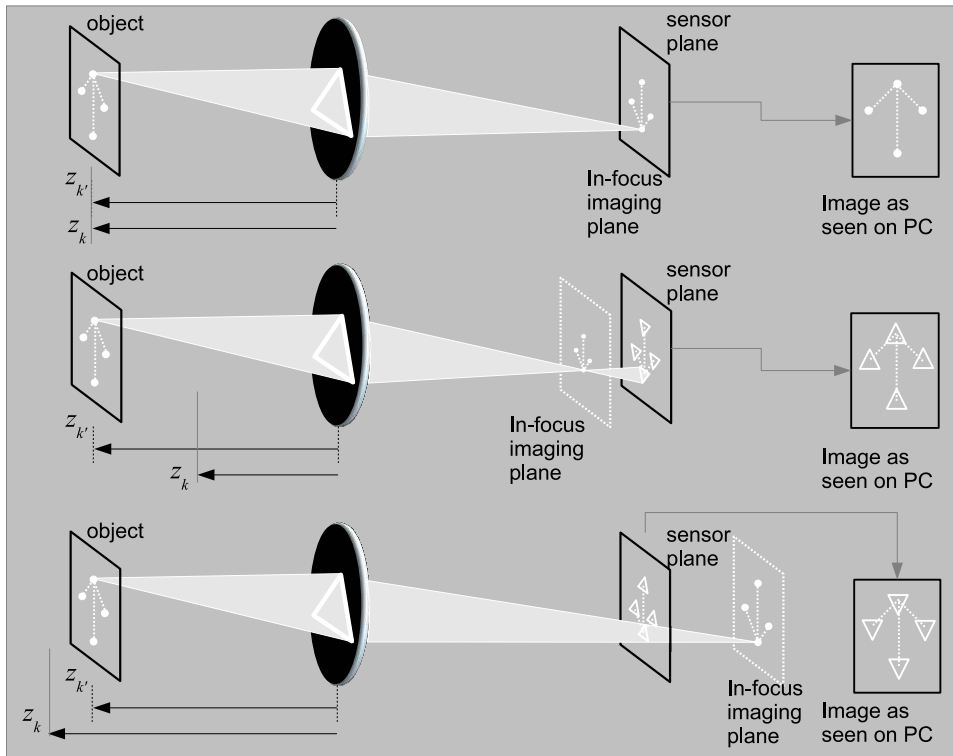
Figure 4.3: Hollow triangular pupil. (Top) When the focusing distance z_k equals the object plane distance $z_{k'}$, the arrow along with four points in it are imaged in-focus in the camera sensor (a point is imaged as a point, *i.e.* the PSF $g_{kk} = \delta$ function). (Middle) distance to the object plane $z_{k'} > z_k$, then the object is imaged out-of-focus in the sensor, the PSF resembles the inverted shape of the pupil (every point is imaged as an inverted triangle on the sensor). (Bottom) distance to the object plane $z_{k'} < z_k$, then the object is imaged out-of-focus in the sensor but now the PSF resembles the shape of the pupil without inversion (every point is imaged as a triangle -with the same orientation as the pupil- on the sensor).

should take this inversion into account:

$$g_{kk'}(x,y) = \begin{cases} \dfrac{p_{kk'}(x,y)}{\iint p_{kk'}(x,y)dxdy}, & z_{k'} \geq z_k \\[3mm] \dfrac{p_{kk'}(-x,-y)}{\iint p_{kk'}(x,y)dxdy}, & z_{k'} < z_k \end{cases} \tag{4.14}$$

(Note that for a center-symmetric pupil $p_{kk'}(-x,-y) = p_{kk'}(x,y)$ and Equation (4.14) reduces to Equation (4.13)).

4.2 Results

To illustrate the effects of the PSFs simulated through Equation (4.14), we acquired an optimal stack of a scene with a small *circuit* as shown in *Visualization acquired circuit optimal stack*. Figure 4.4 shows six of the thirty images in the acquired stack.

As examples of refocusing with reshaped apertures, we considered the reconstruction of images using Equation (4.7) as they would have been acquired through differently shaped apertures. Figure 4.5 shows the refocused scene through different pupils for the same focusing distance $z_k = 55$mm. The considered reshaped pupils were an annular pupil with external radius $R/2$ and internal radius $2/5R$ (Figure 4.5(b)), a hollow triangle inscribed in a circle of radius $R/2$, with the internal triangle being 0.8 of the outer triangle (Figure 4.5(c)), circular pupil with radius $R/2$ (Figure 4.5(d)) and $R/10$ (Figure 4.5(e)) (we can observe the increment in DOF range with the reduction in pupil size) and a pinhole pupil (Figure 4.5(f)) which recovers an extended-depth-of-field, *i.e.* reconstructs the all-in-focus image.

Finally, in Figure 4.6, the reconstructed images for the same focusing distances in Figure 4.4 are shown as acquired through the hollow triangle pupil. Note inversion of the hollow triangle (defocused PSF) at each side of the in-focus plane.

Synthetic aperture reshaping can contribute to a better understanding of the effects of a particular aperture, on numerous disciplines such as coded aperture design and asymmetrical apertures for focal side detection, without the need for any physical change in the optical system.
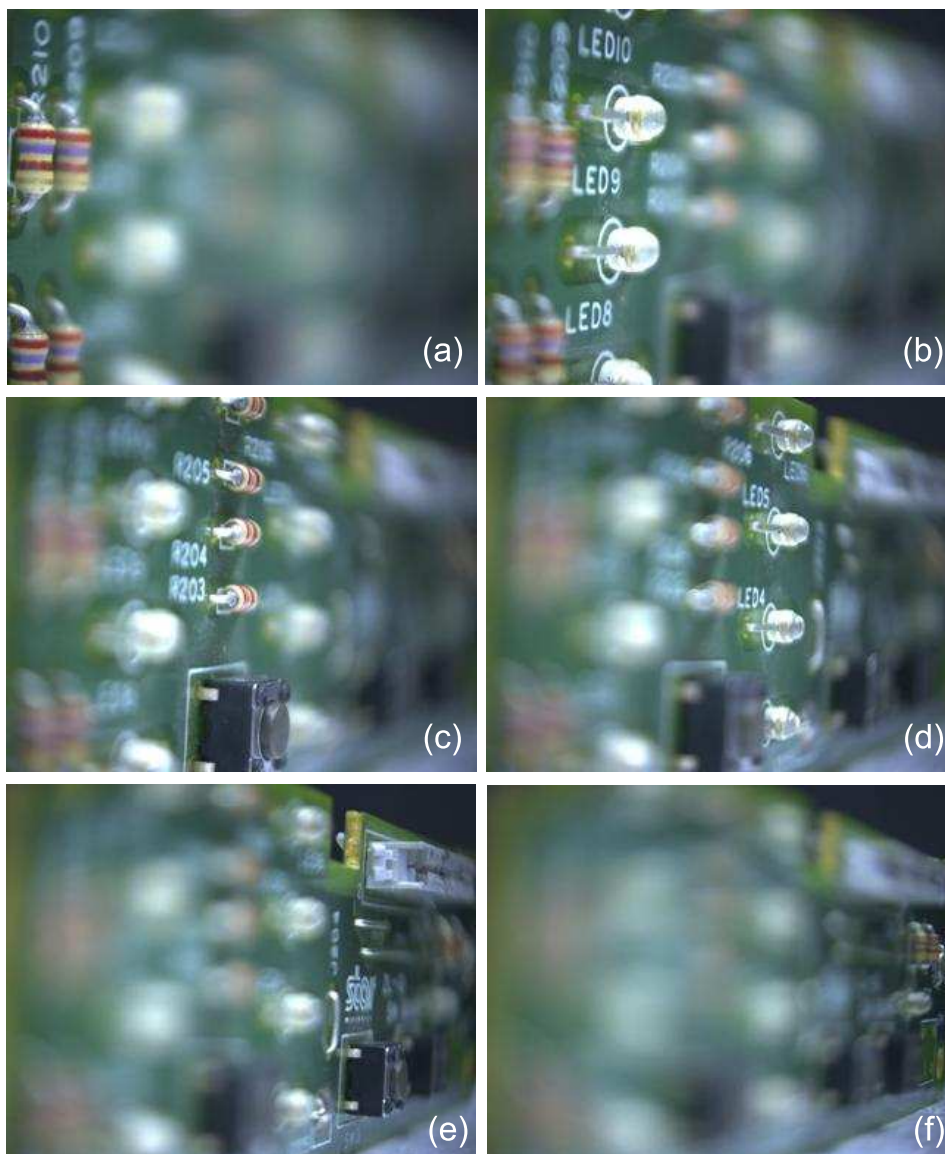
Figure 4.4: Six out of thirty images of the acquired multi-focus image stack with system focusing at (a) 55, (b) 71, (c) 84, (d) 98, (e) 117 and (f) 165mm.
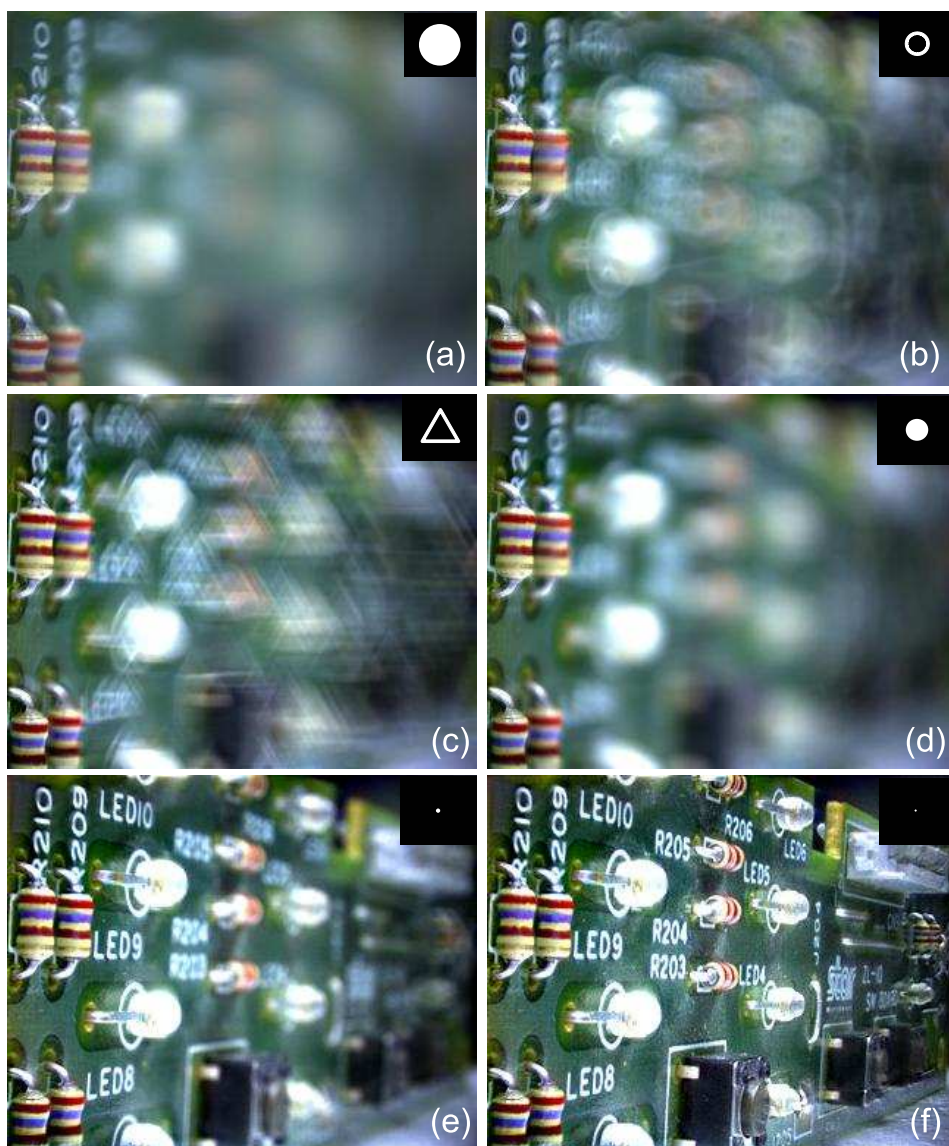
Figure 4.5: (a) Original image acquired with the system focused at 55mm with a circular pupil of radius R. Post-acquisition refocusing with reshaped aperture (the synthetized pupil is shown in the inset); (b) annular pupil with radius $R/2$; (c) hollow triangle; (d) circular pupil with radius $R/2$; (e) circular pupil with radius $R/10$ (partially extended-depth-of-field); and (f) centered pinhole (all-in-focus image).

Figure 4.6: Six images of the refocused multi-focus image stack with a triangular pupil (asymmetric pupil). System focusing at (a) 55, (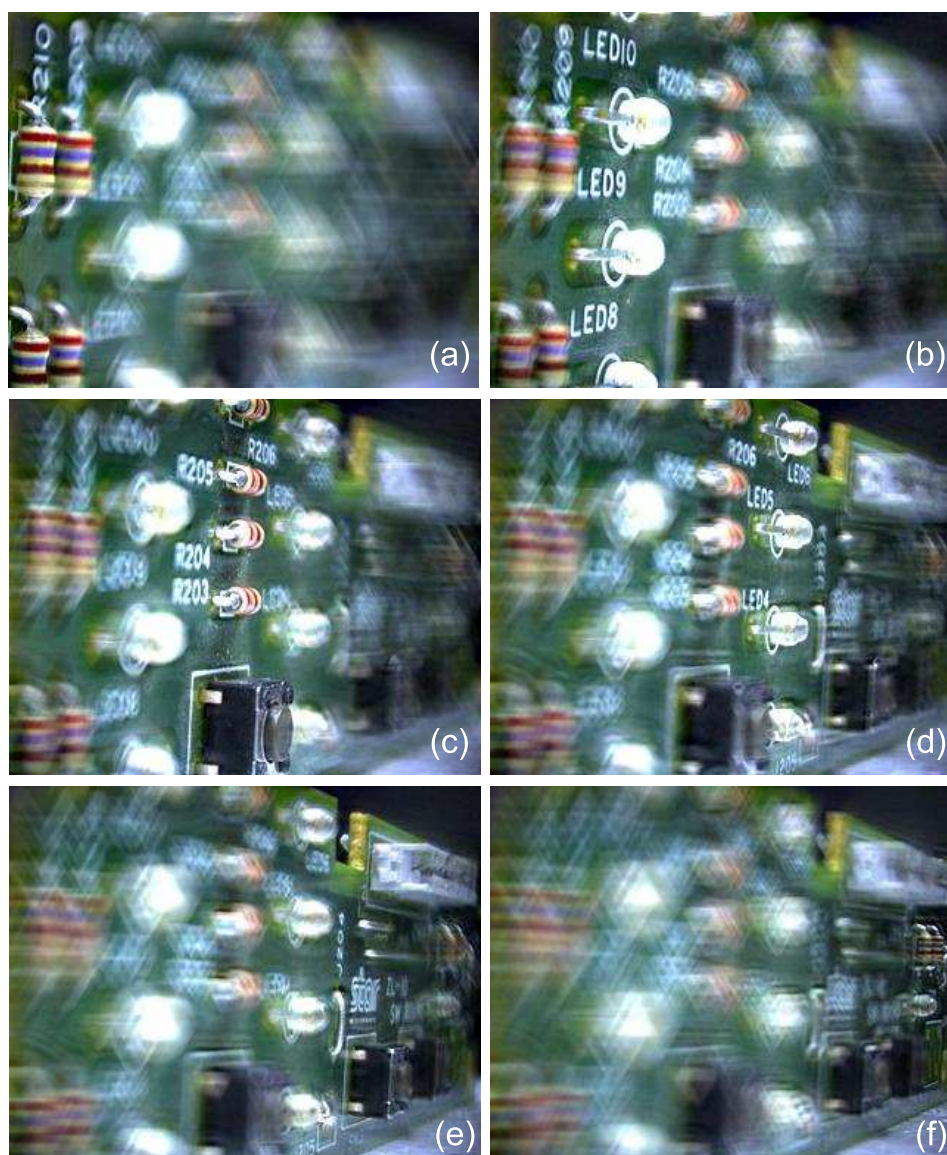b) 71, (c) 84, (d) 98 , (e) 117 and (f) 165mm. Note inversion of the projection of the pupil at each side of the in-focus plane.

Chapter 5

Free viewpoint image synthesis

Cameras with a narrow aperture have a large depth-of-field, so both the background and the foreground appear in-focus at the same time. However, real scenes exhibit occlusions between objects and then the background will only be partially visible because of the occlusion. Images generated with a finite aperture contain information that is lost in the ideal projection through a pinhole, since defocus allows more of the scene to be visible than would be seen through an infinitesimal pinhole, reaching information that is occluded. This is illustrated in Figure 5.1 where a lens on a finite aperture defocuses the foreground, allowing to see the background behind the occluding boundaries. The extent of this phenomenon depends on the aperture, the focal length, the size of the occluding object and its relative distance to the background [Favaro and Soatto, 2003, Favaro and Soatto, 2007]. The same situation occurs when considering a set of differently focused images. The multi-focus stack has information of the background occluded objects which allows the synthesis of different perspective views within some limits.

In recent years, some methods for viewpoint image synthesis from a multi-focus image stack have been proposed. Levin and Durand [Levin and Durand, 2010] synthesized perspective images from a multi-focus or focal stack. The images of a given multi-focus stack are shifted according to their disparity, then they are averaged and this result is deconvolved with a depth-invariant point-spread function (PSF) corresponding to the viewpoint direction. Mousnier et al. [Mousnier et al., 2015], on the other hand, estimated the depth map of a scene from camera calibration and then created a tomographic reconstruction of the epipolar images by back-projection to finally render perspective shifts (shear viewpoints).

The purpose of this chapter is to present a method for post-capture perspective shifts reconstruction in the x, y, and z directions from an optimal multi-focus image stack [Alonso et al., 2016]. We will also see how our method outperforms state-

Figure 5.1: Occluded portions of the background scene are not visible using a pinhole camera (left). Using a finite aperture lens, however, allows one to read the legend "Accommodation will set you free" behind the bars (right). From [Favaro and Soatto, 2003].

of-the-art schemes and unlike depth-invariant approaches, deals well with strongly defocused image stacks. Our method also works without the segmentation of the focused regions or retrieving the depth map, which could introduce inaccuracies into the reconstruction.

5.1 Novel viewpoint synthesis in Fourier domain

In order to implement the synthesis of novel viewpoints, a set of N images is acquired by varying the focusing distances of the optical system according to Equation (2.13). Image i_k ($k = 1, ..., N$) is taken with the system focused at axial distance $z = z_k$ (Figure 5.2). Figure 5.3 shows four out of the twenty-five images acquired through the system described in Chapter 2. (see *Visualization acquired cat stack*). The image formation model follows Equation (4.1):

$$i_k(x, y) = \sum_{k'=1}^{N} h_{kk'}(x, y) * f_{k'}(x, y),$$ (5.1)

with $h_{kk'}(x, y)$ given by Equation (3.13):

$$h_{kk'}(x, y) = \frac{1}{\pi r_{kk'}^2} circ\left(\frac{\sqrt{x^2 + y^2}}{r_{kk'}}\right),$$ (5.2)
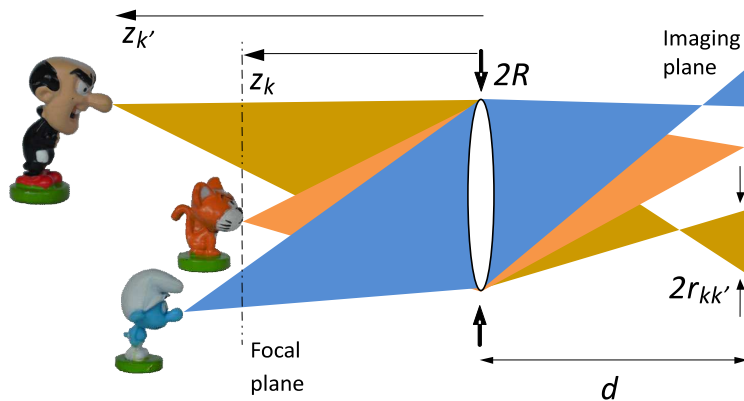
Figure 5.2: Optical system focusing at z_k with a circular aperture. A point in the cat is seen, in-focus, as a point in the imaging plane, while a point in the other plane at a distance $z_{k'}$ is seen out-of-focus as a filled circle of radius $r_{kk'}$ in the imaging plane.

and defocus radius given by Equation (3.21)

$$r_{kk'} = Rd \left| \frac{1}{z_k} - \frac{1}{z_{k'}} \right|. \tag{5.3}$$

where R and d are those indicated in Figure 5.2.

In Fourier domain, image model Equation (5.1) can be written in vector form as

$$\vec{I}(u,v) = H(u,v)\vec{F}(u,v), \tag{5.4}$$

with N-element column vectors $\vec{I}$, $\vec{F}$ and $N \times N$ symmetric matrix H given by Equation (3.29). As we saw before, the Moore-Penrose pseudo-inverse provides the set of vectors that minimize the Euclidean norm $\left\| H(u,v)\vec{F}(u,v) - \vec{I}(u,v) \right\|$ in the least squares sense. Thus, the minimum norm vector is given by

$$\vec{F}(u,v) = H^\dagger(u,v)\vec{I}(u,v). \tag{5.5}$$

When $H(u,v)$ is invertible, then $H^\dagger(u,v) = H^{-1}(u,v)$ and Equation (5.5) reduces to $\vec{F}(u,v) = H^{-1}(u,v)\vec{I}(u,v)$.

We are interested now in those linear combinations of the elements of the minimum norm vectors in Equation (5.5) that lead to the reconstruction of novel viewpoints, without depth-map estimation or segmentation of the in-focus regions (focus slices) f_k.
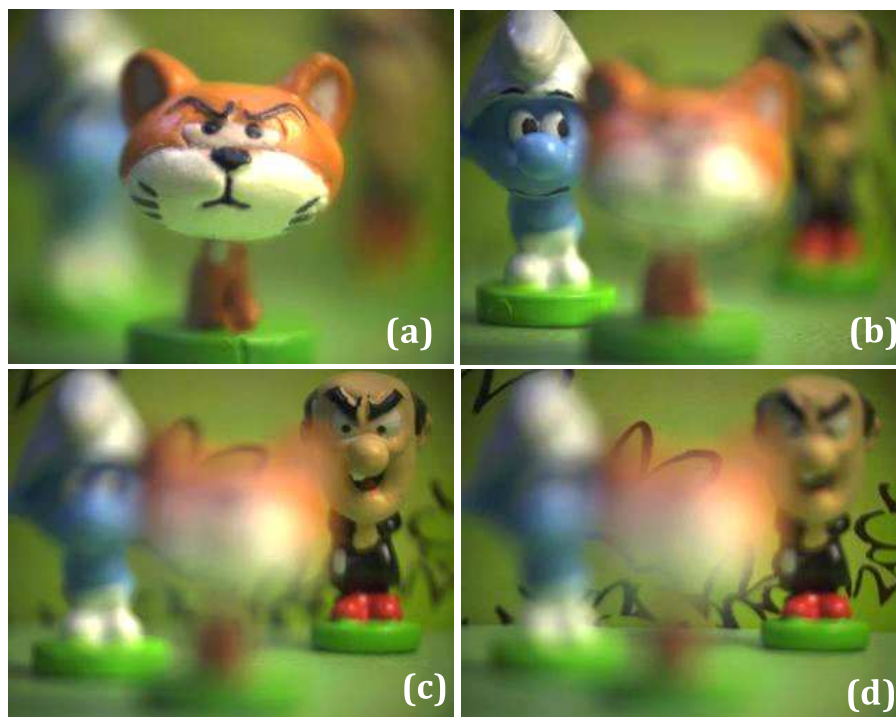
Figure 5.3: Four out of twenty-five images of the stack (*Visualization acquired cat stack*). System focused at; (a) 64; (b) 95; (c) 125, and; (d) 165mm.

5.2 Perspective shifts in horizontal and vertical directions
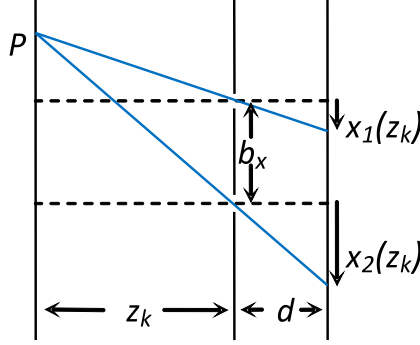
Figure 5.4: Same point P, as seen through a pinhole camera in two different positions separated a distance b_x . Disparity is given by $x_2 - x_1 = \frac{d}{z_k}b_x$.

Let us consider the all-in-focus reconstruction, $s_{b_x,b_y}(x,y)$ from an arbitrary viewpoint of the scene. This perspective shift is generated by simulating the displacement of a pinhole camera in the (x,y) plane (see Figure 5.4). Each focus slice $f_k(x,y)$ should be shifted in an amount according to the disparity associated with the focusing distance z_k and the virtual baseline displacement vector (b_x, b_y) of the camera [Park et al., 2008],

$$s_{b_x,b_y}(x,y) = \sum_{k=1}^{N} f_k\left(x - \frac{d}{z_k}b_x, y - \frac{d}{z_k}b_y\right). \tag{5.6}$$

Fourier transform shift theorem states that translation in the space domain introduces a linear phase shift in the frequency domain [Goodman, 1996]):

Shift theorem. If $\mathcal{F}\{g(x,y)\} = G(u,v)$, then

$$\mathcal{F}\{g(x-a,y-b)\} = G(u,v)e^{-j2\pi(ua+vb)}. \tag{5.7}$$

Thus, by using Equation (5.5) for $\vec{F}(u,v)$, we obtain the Fourier transform of Equation (5.6):

$$S_{b_x,b_y}(u,v) = \sum_{k=1}^{N} e^{-j\frac{2\pi d}{z_k}(b_x u + b_y v)} \left(H^{\dagger}(u,v)\vec{I}(u,v)\right)_k. \tag{5.8}$$

Finally, we reconstruct the new scene perspective as $s_{b_x,b_y} = \mathcal{F}^{-1}\{S_{b_x,b_y}\}$. In particular, $s_{0,0}$ recovers the all-in-focus image reconstruction (see Figure 5.5(a)) obtained in Chapter 3 and considered a particular case (pinhole) of aperture reshaping in Chapter 4. Figures 5.5(b)-5.5(e) show the result of displacing the pinhole camera to the left, right, up, and down, respectively.

5.3 Perspective shifts in axial direction

A pinhole camera displacement b_z along the z-axis (toward or away from the scene) can also be simulated. In this case, from the image formation theory, the focus slice $f_{z_k-b_z}$, as seen under displacement b_z, is related to f_k by:

$$f_{z_k-b_z}(x,y) = \left(\frac{z_k - b_z}{z_k}\right)^2 f_k\left(\frac{z_k - b_z}{z_k}x, \frac{z_k - b_z}{z_k}y\right),\qquad(5.9)$$

where the scale factor $\left|\frac{z_k-b_z}{z_k}\right|$ is given by the change in magnification resulting from the virtual axial displacement (see Fig. 5.6). Then, a new viewpoint from the b_z displacement is given by:

$$s_{b_z}(x,y) = \sum_{k=1}^{N}\left(\frac{z_k - b_z}{z_k}\right)^2 f_k\left(\frac{z_k - b_z}{z_k}x, \frac{z_k - b_z}{z_k}y\right).\qquad(5.10)$$

From the Fourier transform similarity theorem [Goodman, 1996], which states that a stretch of coordinates in space domain results in a contraction in frequency domain:

   Similarity theorem. If $\mathcal{F}\{g(x,y)\} = G(u,v)$, then

$$\mathcal{F}\{g(ax,by)\} = \frac{1}{|ab|}G(\frac{u}{a},\frac{v}{b}),\qquad(5.11)$$

   we obtain:

$$S_{b_z}(u,v) = \sum_{k=1}^{N} F_k\left(\frac{z_k}{z_k - b_z}u, \frac{z_k}{z_k - b_z}v\right).\qquad(5.12)$$

Finally, from Equation (5.5) we obtain S_{b_z}:

$$S_{b_z}(u,v) = \sum_{k=1}^{N}\left(H^{\dagger}\left(\frac{z_k}{z_k - b_z}u, \frac{z_k}{z_k - b_z}v\right)\vec{I}\left(\frac{z_k}{z_k - b_z}u, \frac{z_k}{z_k - b_z}v\right)\right)_k.\qquad(5.13)$$

and reconstruct the new scene perspective as $s_{b_z} = \mathcal{F}^{-1}\{S_{b_z}\}$. The results for the backward and forward displacements are shown in Figs. 5.5(f) and 5.5(g), respectively. A continuous change of viewpoints in the (x,y,z) directions can be seen in
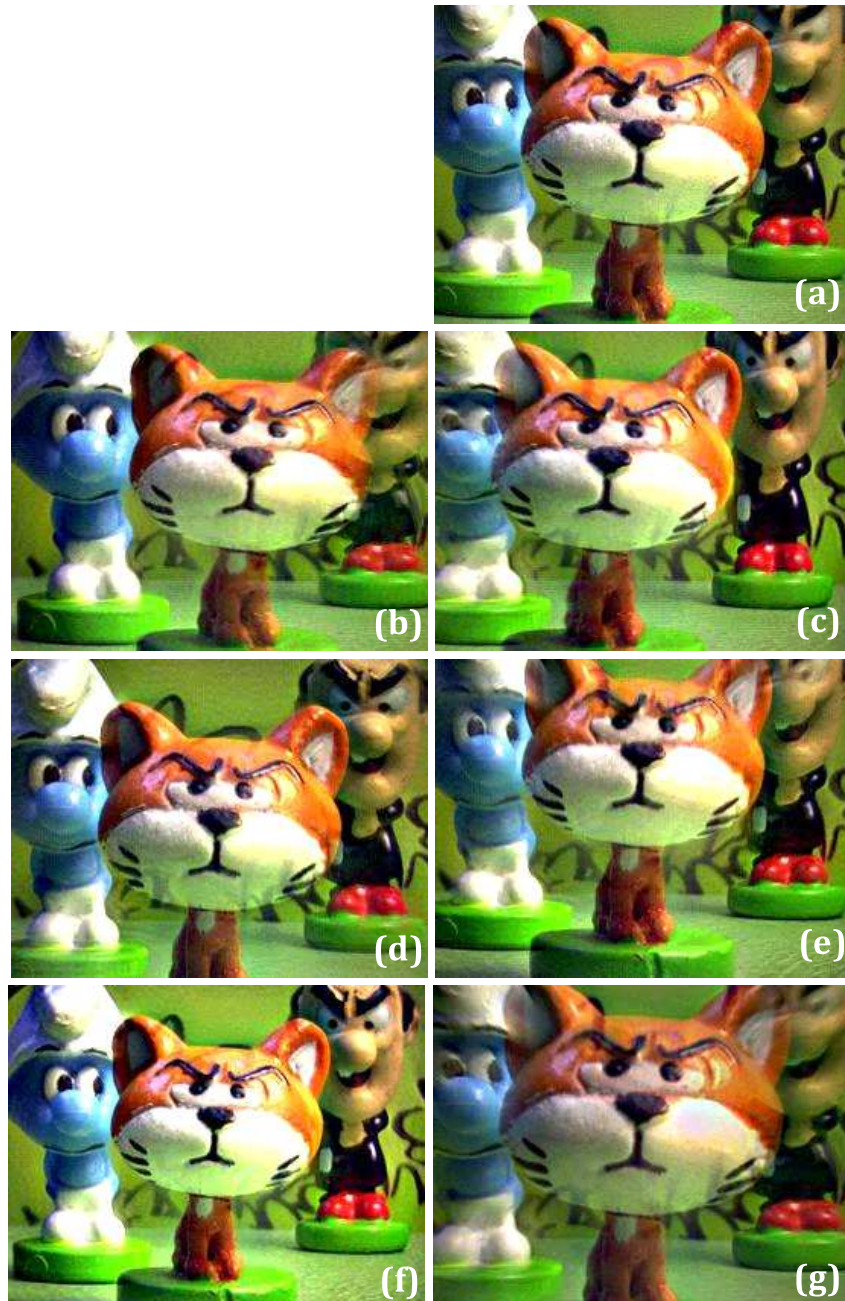
Figure 5.5: Novel viewpoints (*Visualization cat novel viewpoints*) through a simulated pinhole placed at (b_x, b_y, b_z) equal to; (a) $(0,0,0)$; (b) $(2,0,0)$; (c) $(-2,0,0)$; (d) $(0,2,0)$; (e) $(0,-2,0)$; (f) $(0,0,-10)$, and; (g) $(0,0,10)$mm.
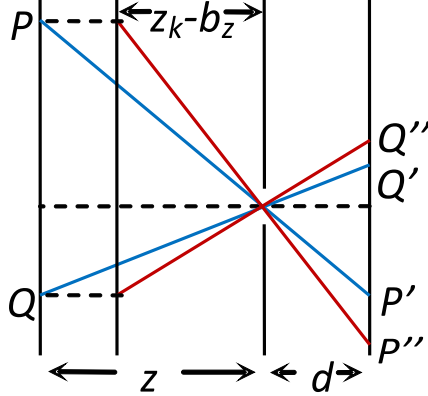
Figure 5.6: Segment PQ, as seen through a pinhole camera at distance z_k ($P'Q'$) and at distance $z_k - b_z$ ($P''Q''$). The scale change factor is given by $\frac{P''Q''}{P'Q'} = \frac{z_k}{z_k - b_z}$.

Visualization cat novel viewpoints.

5.4   Performance study

5.4.1   Performance comparison

For comparison purposes, similar perspective shifts in the x direction were generated following the methods proposed in [Levin and Durand, 2010] and [Mousnier et al., 2015]. The result from [Levin and Durand, 2010] (see Figure 5.7(a)) led to a blurrier reconstruction in comparison to ours (Figure 5.7(c)). This is probably associated with the process of averaging shifted images and the spatially uniform depth-invariant deconvolution, which might not work well for severely defocused stacks. On the other hand, the result from [Mousnier et al., 2015] (Figure 5.7(b)) inherits some inaccuracies from the depth-map estimation and region growing implicit in the method (note, for example, the patches in the nose, head, and left ear of the cat).

Besides visual comparisons, quantitative comparisons for a synthetic multi-focus image stack were performed. Figure 5.8(a) shows a synthetic 3D scene constructed from four plane objects from the image *peppers* available in Matlab. Each object was placed at a given distance from the camera and the multi-focus image stack (Figures 5.8(b)-5.8(e)) was constructed considering the same parameters R, d, and p as those for the real optical system (see Chapter 2).

Unlike with the real multi-focus stack, we can construct a *ground truth* (GT)

Figure 5.7: Perspective shift result comparison for a similar viewpoint (a) Levin and Durand [Levin and Durand, 2010] for $u = -30$ ($b_x = -1$mm), (b) Mousnier et al. [Mousnier et al., 2015] for $u = -30$ ($b_x = -1$mm), and (c) proposed method with $b_x = -1$mm.

reference for each perspective shift of interest. Figures 5.9(a)-5.9(c) show the GTs for the scene as viewed from a shifted pinhole camera (to the left, center, and right, respectively). The synthetic stack of Figs. 5.8(b)-5.8(e) is used to render the same viewpoints using the work from [Levin and Durand, 2010], [Mousnier et al., 2015] and the proposed method. The visual results for Figures 5.9(d)-5.9(f), show blurry images (as for the real case Figure 5.7), which are probably associated with the process of averaging the shifted images and the spatially uniform depth-invariant deconvolution considered in [Levin and Durand, 2010], which might not work well for severely defocused multi-focus image stacks.

A visual inspection of Figures 5.9(g)-5.9(i) shows visual artifacts in the images generated using the method of [Mousnier et al., 2015] due to errors in the depth estimation process (based on the detection of strong gradients) and region growing (where strong gradients have been detected) for a severely defocused stack. On the other hand, the images in Figs. 5.9(j)-5.9(l) are the result of the proposed method. It can be seen that our method achieves better results due to the consideration of a depth-variant PSF and deals better with strongly defocused image stacks. Table 5.1 shows a quantitative comparison (of luminances) between the different methods. The values labeled in bold indicate the best performance for each perspective. The lower mean square error (MSE, see Equation (3.36)) value indicates a result closer to the GT reference, showing that our method outperforms state-of-the-art schemes in all the analyzed cases.
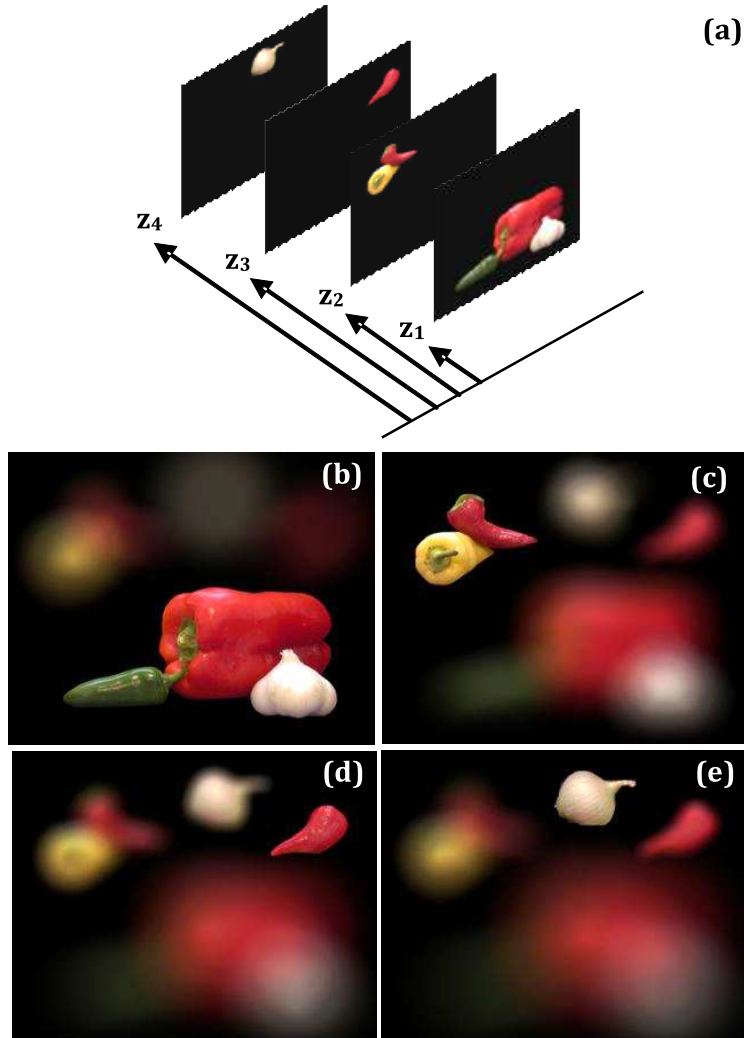
Figure 5.8: Synthetic defocused images generated from *peppers*. (a) 3D synthetic scene. System focused at; (b) $z1 = 60$mm; (c) $z_2 = 90$mm; (d) $z_3 = 120$mm, and; (e) $z_4 = 150$mm.

Table 5.1: MSE Value Comparison Between Different Methods.

b_x(mm)	Levin-Durand [Levin and Durand, 2010]	Mousnier et al. [Mousnier et al., 2015]	Proposed method
+1	712.0578	135.8557	**17.8339**
0	672.5608	67.9387	**13.4741**
-1	712.9319	186.0165	**16.6159**
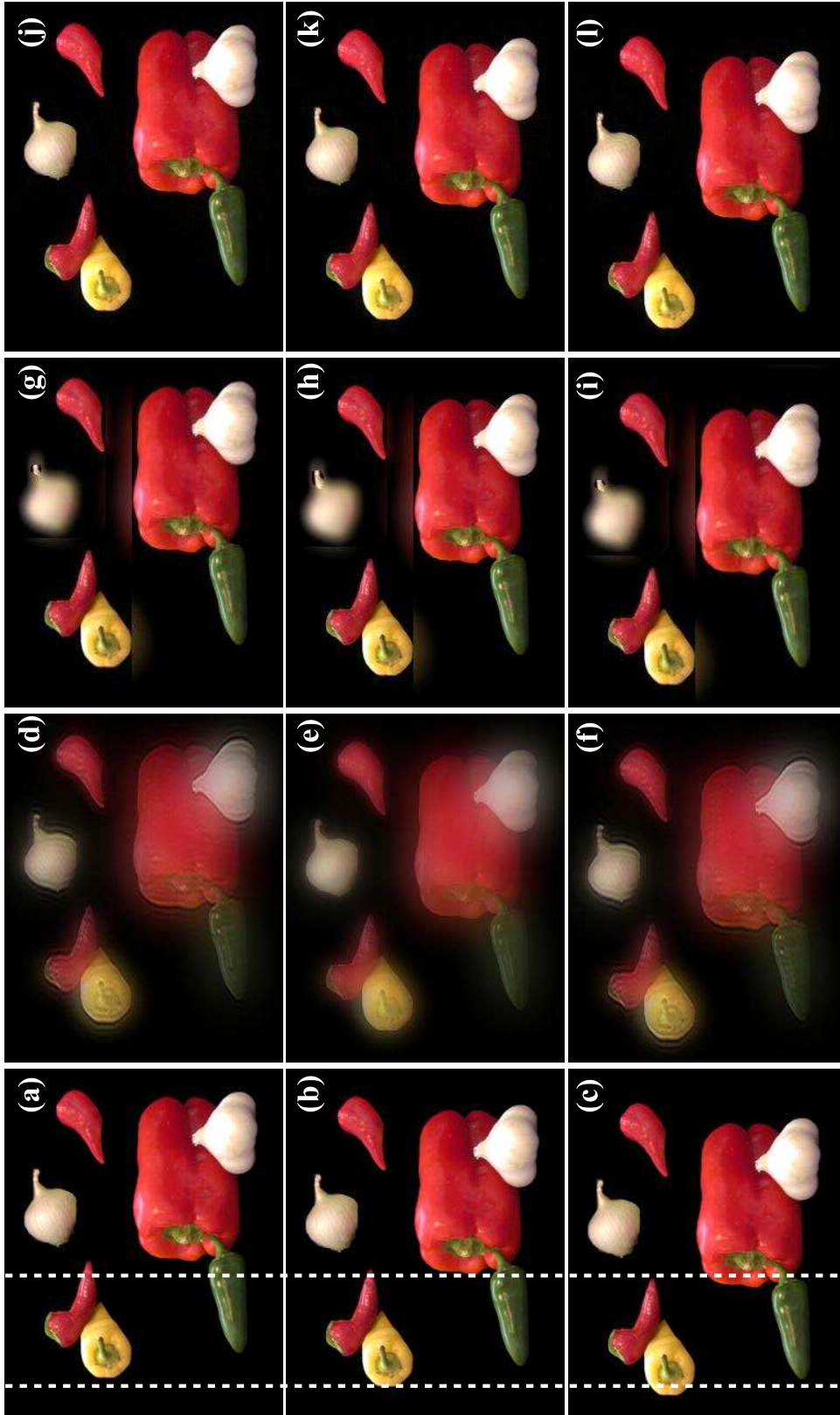
Figure 5.9: Perspective shift comparison for viewpoints corresponding to $b_x = 1$, 0, and -1mm. (a)-(c) Ground truth; (d)-(f) Levin and Durand, for $u = 30, 0$, and -30; (g)-(i) Mousnier et al., for $u = 30, 0$, and -30; (j)-(l) proposed method. Two vertical dotted lines aligned to the left edge of two objects (at different depths in the scene) in the frontal view, (b), allow us to see clearly the different shifts in those objects in (a) and (c) according to the disparity corresponding with its focusing distance.

5.4.2   Perspective shifts limits

It is worth to mention that an inherent problem of the view synthesis is due to the fact that areas, which are occluded in the original view, might become visible (exposed or disoccluded) in the new shifted perspective [Fehn et al., 2006]. In our approach this is not a significant issue as can be observed for example in Figure 5.5(b). When the virtual pinhole camera moves to the left, the cat allows to see more background on the left of the image, without significant artifacts. However, a slight see through effect is observed on the right of the image at occlusion boundaries, where areas which are disoccluded in the original view might become occluded. This effect is expected since the convolution image formation model does not consider occlusion.

In order to determine the limit case for point-of-view displacement we will consider the configuration depicted in Figure 5.10. The 3D scene consists of a blue disk at distance z_k from the camera occluding a white plane at distance $z_{k'} > z_k$. For the system focused at $z_{k'}$ (Figure 5.10(b)) the blurred part of the image comes from the convolution of the blue disk with a circle of radius $r_{kk'}$ given by Equation (5.3):

$$r_{kk'} = Rd \left(\frac{1}{z_k} - \frac{1}{z_{k'}} \right).$$

$$(5.14)$$

where we omitted the modulus since $z_{k'} > z_k$.

This blurring is responsible for a see-through effect around the edge of the object but leaves the innermost part of the blue disk untouched such that the region enclosed in the red dotted circle is not visible.

Let us now consider a horizontal displacement b_x of the viewpoint of the all-in-focus scene (Figure 5.10(c)). According to Section 5.2, the displacement for the disk at z_k is:

$$\delta_k = \frac{db_x}{z_k},$$

$$(5.15)$$

while for the plane at $z_{k'}$:

$$\delta_{k'} = \frac{db_x}{z_{k'}}.$$

$$(5.16)$$

The relative displacement between the focus slices is given by:

$$\Delta_{kk'} = \delta_k - \delta_{k'} = db_x \left(\frac{1}{z_k} - \frac{1}{z_{k'}} \right),$$

$$(5.17)$$

This relative displacement should not exceed $r_{kk'}$, since we have absolutely no access to the information of whatever is in the background, inside the red dotted circle:

$$\Delta_{kk'} \leq r_{kk'},$$

$$(5.18)$$

so by substituting Equations (5.14) and (5.17) into Equation (5.18) we arrive to the following condition over the displacement b_x:

$$db_x \left(\frac{1}{z_k} - \frac{1}{z_{k'}} \right) \leq Rd \left(\frac{1}{z_k} - \frac{1}{z_{k'}} \right) \Rightarrow b_x \leq R, \tag{5.19}$$

that is, displacements of the viewpoint cannot exceed the actual size of the original circular aperture. The same result for the limit in the displacement of the viewpoint in Equation (5.19) was obtained in [Schechner and Kiryati, 2000].
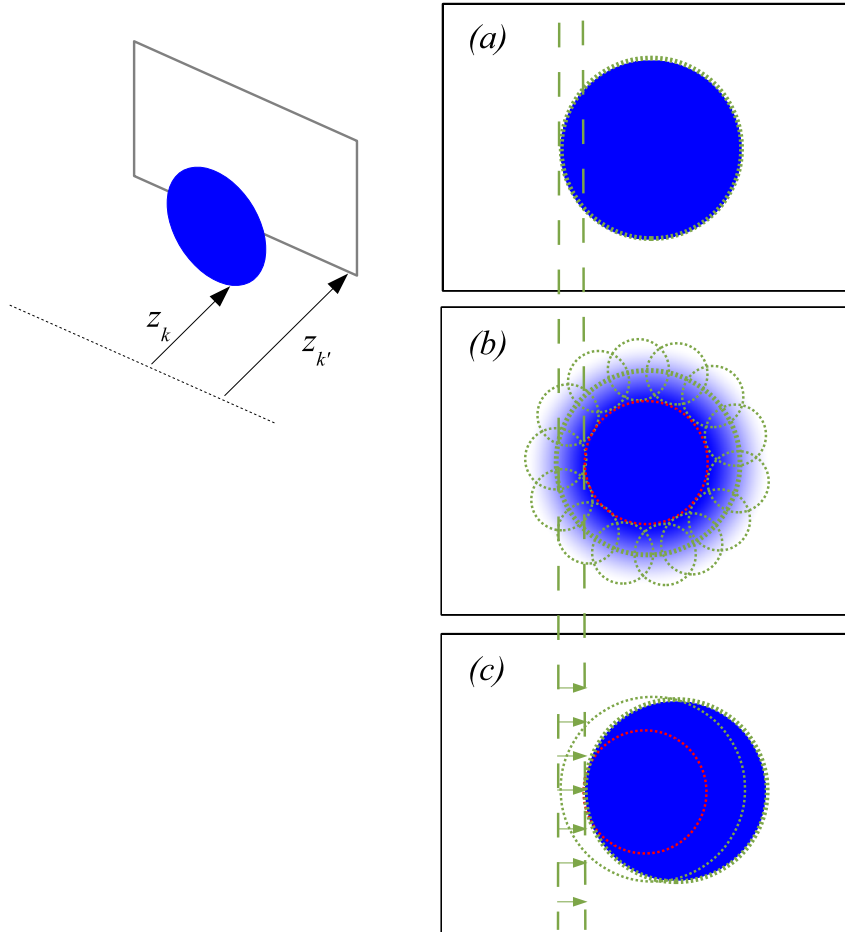
Figure 5.10: Synthetic scene to study the limits for perspective shifts. (Left) 3D synthetic scene configuration, consisting of a blue disk at distance z_k from the camera, in front of a white plane at distance $z_{k'}$ ($z_{k'} > z_k$). (a) Image with the system focused at z_k; (b) System focused at $z_{k'}$ where the blurring circles of radius $r_{kk'}$ (small circles in green dotted line) leave out a totally occluded zone in the background (enclosed by the red dotted circle). (c) Maximum synthesized relative displacement (green arrows) between the circle in the foreground (at z_k) and the plane in the background (at $z_{k'}$) must not exceed $r_{kk'}$.

Chapter 6

Stereoscopic 3D-scene synthesis

6.1   Stereoscopy and binocular vision

Binocular vision is based on the fact that objects in the three-dimensional (3D) surrounding world are perceived from two different perspectives due to the horizontal separation between our left and right eyes. As a result, the left and right images of a 3D scene in our retinas are slightly different. This retinal disparity between the images provides the observer with information about the relative distances and depth structure of three-dimensional objects. Both images are fused by our brain in a process called stereopsis to give us the perception of depth (see Figure 6.1). Stereopsis thus acts as a strong depth cue, particularly at short distances [1]. Stereoscopic vision is one of the most investigated depth perception mechanisms, and a large body of literature exists about it [Howard and Rogers, 1995, Lambooij et al., 2009].

Until the advent of the cinema, the stereoscope was the optical wonder of the age [Howard and Rogers, 1995]. The first stereoscopic display was created by Sir Charles Wheatstone in 1833 (Figure 6.2(a)) and consisted of a mirror device that enabled the viewer to fuse two slightly different views of the same drawing into one stereoscopic image . In 1844, Sir David Brewster further developed the stereoscope (Figure 6.2(b)) by utilizing prismatic lenses and in a short time it became a commercial success . This device was further refined by Oliver Wendell Holmes (Figure 6.2(c)) [Howard and Rogers, 1995, Howard and Rogers, 2012, Surman, 2013]. In 1867 James Clerk Maxwell devised a real image stereoscope for the viewing of mathematical figures (a family of algebraic surfaces known as *the cyclide*) (Fig-

---

[1]Stereoblindness is the inability to perceive stereoscopic depth by fusing (combining) images from the two eyes. It is estimated to affect 3 to 15 percent of the population. The more common causes are conditions like strabismus (the eyes are incapable of looking in the same direction) or amblyopia (the brain does not use both eyes equally, the "good" eye becomes strongly dominant generating significant differences in visual acuity) [Mendiburu, 2012, Fong et al., 2015].
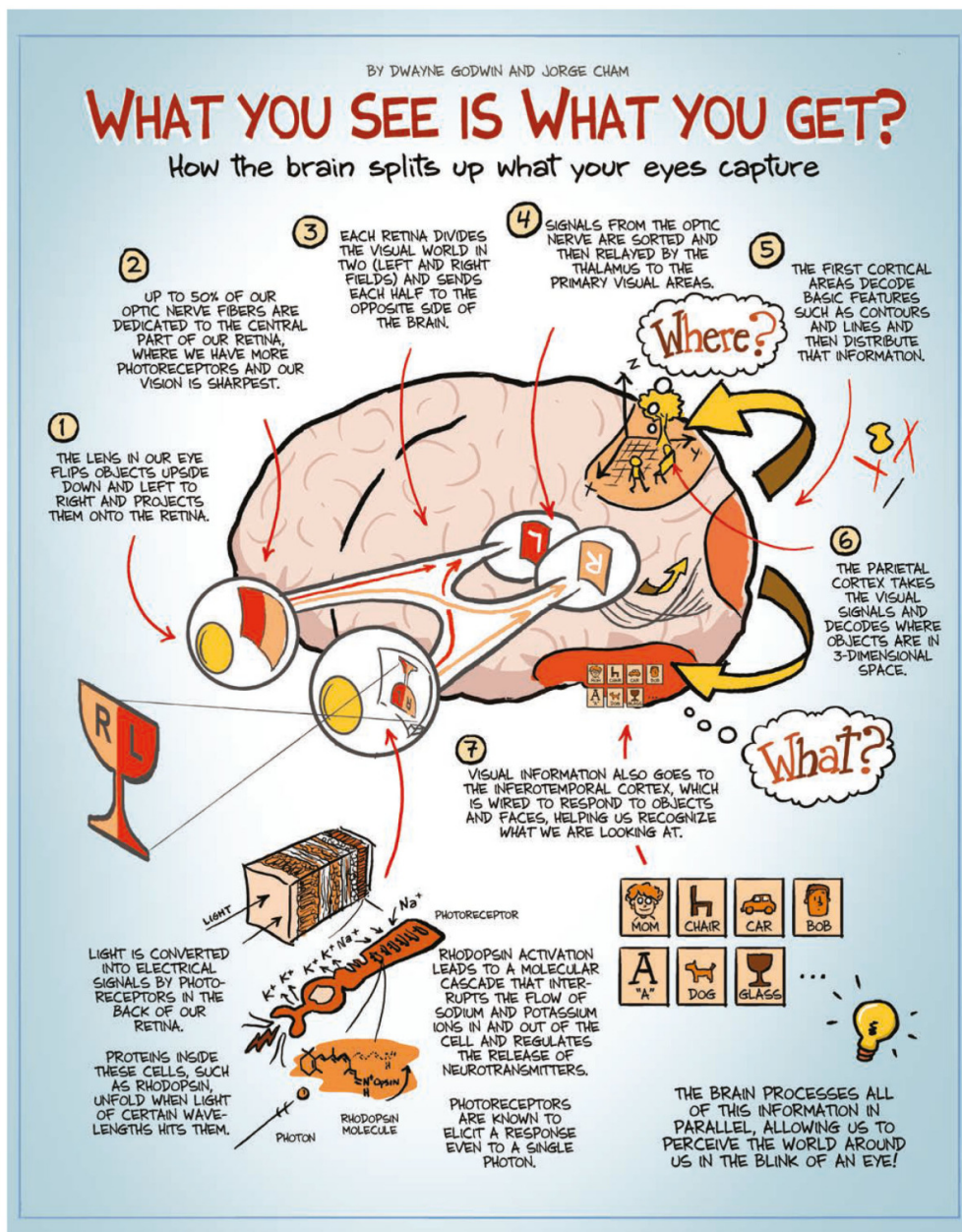
Figure 6.1: Binocular vision: we perceive the world from two slightly different perspectives due to the horizontal separation between our eyes. Retinal disparity gives us information for depth perception in a complex process in our brain known as stereopsis. (Cartoon from Scientific American http://www.scientificamerican.com/article/how-we-see/)

ure 6.2(d)) [Harman and De Regt, 1996, Cat, 2013, Funk, 2012]. Recently, in 2014, Google released the *Google Cardboard* (Figure 6.2(e)), a cardboard stereo-scope for smartphones where apps on the smartphone substitute for stereo cards (https://vr.google.com/cardboard/). In 2015 Samsung in collaboration with Oculus, released the *Samsung Gear VR* (Figure 6.2(f)), a mobile virtual reality headset (http://www.samsung.com/global/galaxy/gear-vr/). These two recent stereoscopes take advantage of modern technology. While Google Cardboard uses the sensors in the smartphone (accelerometer, gyroscope, etc.) to adapt the viewing content, the Samsung Gear VR, uses the sensors at the head mounted display, offering a better VR experience and head tracking. Although considerable progress has been made since the early 1900's in developing stereoscopic methods and technologies, the underlying principle of presenting two different images, corresponding to the different perspectives of the right and left eye, has essentially remained unchanged.

Steroscopic displays can be categorized based on the technique used to channel the right and left images to the appropriate eyes. Those displays that do not require a viewing aid (e.g., glasses) to separate the right and left eye images are known as *autostereoscopic. Direction-multiplexed autostereoscopic displays* apply optical principles such as diffraction, refraction, reflection and occlusion to direct the light from the different perspective views to the appropriate eye. Autostereoscopic techniques based on *parallax barriers* and *lenticular arrays* are still popular today [Iizuka, 2013]. *Parallax barrier* displays (see Figure 6.3(b)) are based on the principle of occlusion, that is, part of the image is hidden from one eye but visible to the other. At the right viewing distance and angle, one eye can only see the appropriate view, as the other view is occluded by the barrier effect of vertical slits. In *lenticular systems* (see Figure 6.3(d)) an array of vertically oriented cylindrical lenses is placed in front of columns of pixels alternately representing parts of the left and right eye view. Through refraction, the light of each image point is emitted in a specific direction in the horizontal plane. This technique is well-known from 3D picture postcards and photographic prints. Both lenticular and parallax barrier systems require the precise alignment of the picture splitter (vertical stripes or lenses) with the vertical left-right image strips [Pastoor and Wöpking, 1997, Lambooij et al., 2009].

Displays that require viewing glasses (see Figure 6.3), can be *time-parallel*, with both left and right eye views appearing simultaneously, or *time-multiplexed*, where the left and right eye views are shown in rapid alternation and synchronized with a liquid crystal (LC) shuttering system which opens in turn for one eye, while occluding the other (see Figure 6.3(a)). For time-parallel stereoscopic displays several multiplexing methods have been used, based on either color, polarization or location. In color multiplexed, or *anaglyph* displays, the left and right eye images are filtered
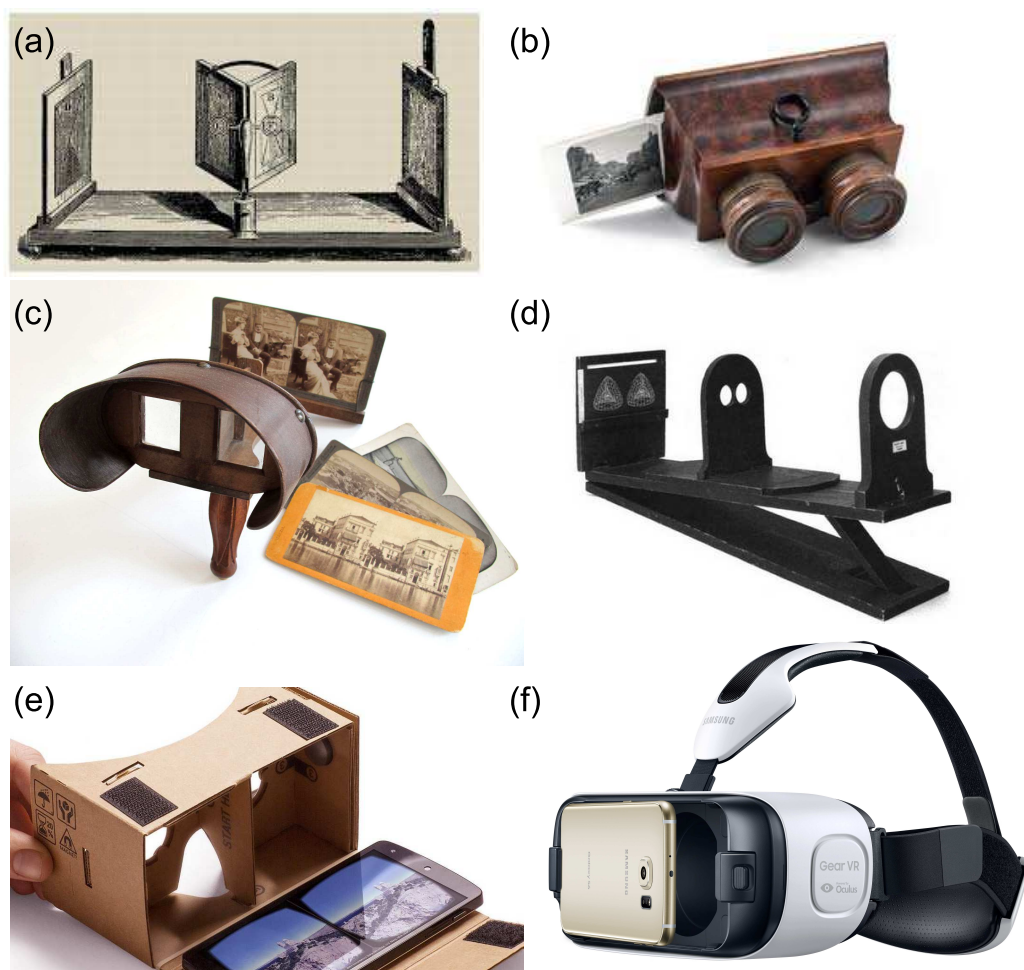
Figure 6.2: (a) Wheatstone mirror stereoscope (1838), (b) Brewster portable stereoscope (1849), (c) Holmes stereoscope (1861) (https://en.wikipedia.org/wiki/Stereoscope), (d) Maxwell real image stereoscope (1867), (e) Google Cardboard stereoscope (2014) (https://vr.google.com/cardboard/) and (f) Samsung Gear VR (2015)

with complementary colors (e.g., red and cyan). The observer is required to wear color-filter glasses to separate the images. This well-known and inexpensive method has been used for stereoscopic cinema (Figure 6.4) and television, and is still popular for viewing stereoscopic images in print [Iizuka, 2013]. A serious limitation of this method is that color information is lost since it is used as a selection mechanism and they may produce retinal rivalry[1]. *Polarization-multiplexed* displays (see Figure 6.3(c)) separate left an right eye images by means of polarized light (*space-multiplexed*). Left and right output channels (monitors or projectors) are covered by orthogonally oriented filters, using either linear or circular polarization. The observer needs to wear polarized glasses to separate the different views again. This is the system most commonly used in stereoscopic cinemas today. The most significant drawback of this kind of system is the loss of light output from using the polarizing filters [Pastoor and Wöpking, 1997].
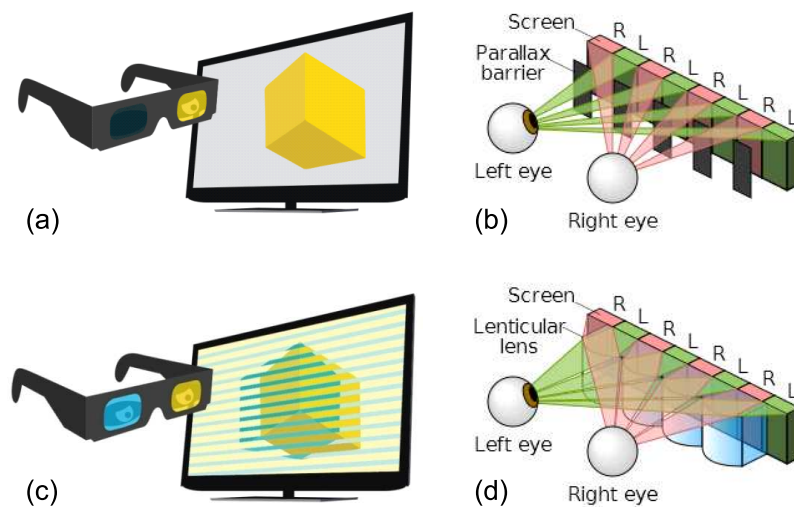
Figure 6.3: (a) Active 3D shutter (time-multiplexed), (b) Parallax barrier (occlusion), (c) Polarized-multiplexed and (d) lenticular lens array (refraction). (From https://www.locafox.de/magazin/3d-fernseher-aktiv-oder-passiv-3d and https://commons.wikimedia.org/w/index.php?curid=14886721)

Stereoscopic displays have become important for many applications; in the entertainment industry, stereoscopic games and movies results in an increased immersion, enhancing the viewer's experience [Schild et al., 2012, Mendiburu, 2012]. When focusing on medical applications, such as medical diagnosis, pre-operative planning, image-guided precision surgery, or medical training, stereoscopic displays have a

---

[1]Retinal or binocular rivalry is a phenomenon of visual perception in which perception alternates between different images presented to each eye instead of perceiving a fused image.

Figure 6.4: Stereoscopic 3D movie exhibition (from http://www.imagineatlanta.com/projects/).

number of characteristics that make their application to these fields advantageous [Fong et al., 2015, van Beurden et al., 2012, Ferdig et al., 2015]. Surgical systems like *da Vinci* or *Viking 3D/HD* allow surgeons to have a better depth perception and an easier determination of anatomic structures with 3D systems than 2D conventional systems [Bilgen et al., 2013, Kunert et al., 2013].

As mentioned earlier, in spite of the rapid evolving changes in technologies, stereoscopic displays still rely on a pair of stereoscopic images, one image for the left eye and one for the right eye. In the present chapter we introduce a new method to reconstruct a pair of stereoscopic images of a 3D-scene from a multi-focus image stack [Alonso, 2016b], based on the view synthesis previously discussed in Chapter 5 and captured with the optical system described in Chapter 2.

6.2 Stereoscopy and horizontal parallax

The disparity between the left and right views produces a relative displacement of an object viewed along two lines of sight, which is called parallax. It measures the difference in the relative position of the image of a given object between the right eye image and the left eye image. Both perspectives of the same 3D scene are fused by the brain to give the perception of depth. When an image point occupies the same position on the screen for both the left and right eyes it has what is referred to as zero disparity or zero parallax.

In a natural environment, when we look at an object, our eyes are converging and accommodating (focusing) on that object. Looking at something involves the visual motor system, *i.e.* the muscles that control the eye's movements as well as the focal length changes of the lenses of our eyes to focus on the object that we are looking at. Under natural viewing conditions, accommodation and convergence (through *vergence* eye movements [1]) work closely, *i.e.* our eyes converge and focus at the same point of an object, thus bringing the object within the limits of fusion. As convergence involves the visual-motor system, the muscles that control the eye's movements, when looking at an object, both eyes are aiming precisely at that object [Mendiburu, 2012]. When we see a 3D movie or a stereoscopic image, we are accommodating on the screen while we may be converging somewhere else (in-front or behind the screen plane) depending on the perceived depth due to horizontal parallax. This decoupling between accommodation and convergence is not natural for the visual-motor system and may cause discomfort (this is usually referred to as "accommodation/convergence conflict" or simply "A/C conflict"). According to parallax an object could be perceived on the screen plane (Figure 6.5(b)), in-front of it (Figure 6.5(c)) or behind it (Figure 6.5(a)). A disparity of zero means the object is perceived as being on the screen plane. If the left eye's view of an object is shifted left at the screen relative to the right view, then the disparity is uncrossed or positive and the object is perceived as behind the screen. On the other hand, if the left view is shifted to the right, then the disparity is crossed or negative and the object is perceived as in-front of the screen.

6.3 Lateral perspective shifts

The reconstruction of an arbitrary horizontal viewpoint of the scene is accomplished by simulating the displacement of a pinhole camera in the horizontal direction with respect to the center of the original pupil. Let us start by considering the disparity between the images of a given point as seen by the camera sensor of a centered pinhole camera and a pinhole camera displaced to the left. As it can be seen by similar triangles in Figure 6.6, the horizontal disparity is given by:

$$x_{shift} - x_0 = \frac{b_x d}{z},\tag{6.1}$$

(same result as in Figure 5.4).

Then, in the piecewise-plane approximation of the 3D scene, to obtain a shifted

---

[1] Vergence eye movements occur when the eyes move simultaneously inward (convergence) or outward (divergence).
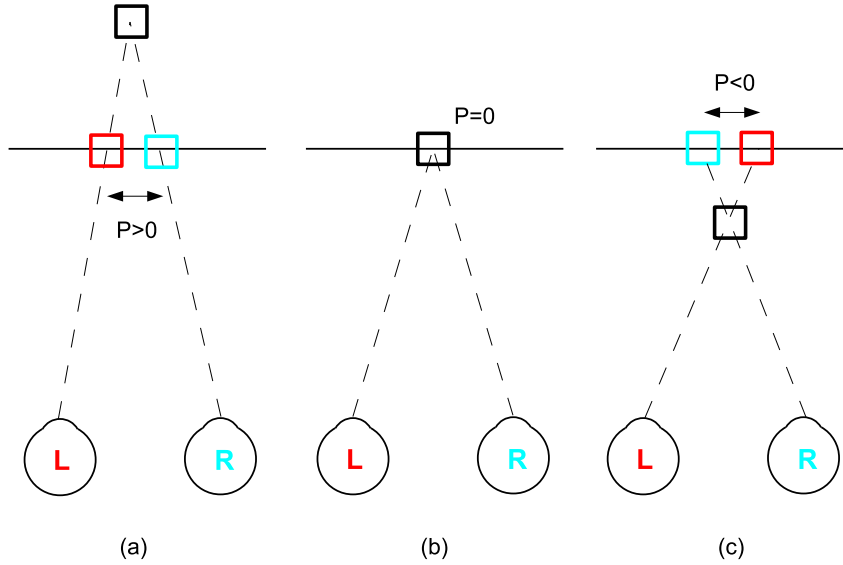
Figure 6.5: Parallax (P) and depth perception. For clarity, left eye image is shown in red while right eye image is shown in cyan. C and A denote convergence and accommodation, respectively. (a) Uncrossed or positive parallax: the object is perceived as behind the screen plane, (b) zero parallax: the object is perceived as in the screen plane and (c) crossed or negative parallax: the object is perceived as in-front of the screen plane (*pop-out of the screen*).
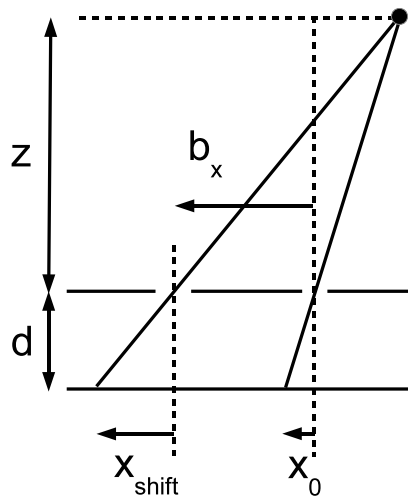
Figure 6.6: Horizontal shift and disparity for a two parallel pinhole camera configuration separated by a baseline b_x. A point in 3D space is imaged at x_0 in the sensor of one pinhole camera (the one in the center of the original pupil) while it is imaged at x_{shift} by the other.

viewpoint $s_{b_x}(x, y)$, each focus slice $f_k(x, y)$ should be shifted in an amount according to the disparity associated with the focusing distance z_k and the baseline displacement (b_x) of the camera (recall Section 5.2),

$$s_{b_x}(x, y) = \sum_{k=1}^{N} f_k \left(x - \frac{d}{z_k} b_x, y \right). \tag{6.2}$$

By means of the Fourier transform shift theorem and following the procedure of Section 5.2, we obtain the Fourier transform of Equation (6.2):

$$S_{b_x}(u, v) = \sum_{k=1}^{N} e^{-j \frac{2\pi d}{z_k}(b_x u)} \left(H^{\dagger}(u, v) \vec{I}(u, v) \right)_k. \tag{6.3}$$

Finally, we obtain the new scene perspective as seen from a pinhole camera, translated b_x to the left of the center of the original circular pupil by inverse Fourier transforming:

$$s_{b_x} = \mathcal{F}^{-1} \{ S_{b_x} \}. \tag{6.4}$$

6.4 Reconstruction and display of a stereoscopic pair of images from a multi-focus stack

6.4.1 Reconstruction of a stereoscopic pair

In the proposed method, x_0 corresponds to the image captured through a centered pinhole (all-in-focus) while x_{shift} corresponds to the left and right images for $b_x = B/2$ and $b_x = -B/2$, respectively (see Section 6.3). This corresponds to a baseline (separation) between the left and right synthetic pinhole cameras of B. The pair of stereoscopic images can then be generated by considering a synthetic stereo-camera formed by a left pinhole camera displaced to the left of the center of the original pupil, $b_x = B/2$, and a right pinhole camera displaced to the right of the center of the original pupil, $b_x = -B/2$.

Then, it is straightforward to reconstruct the left and right views according to Equation (6.3):

$$S_{B/2}(u, v) = \sum_{k=1}^{N} e^{-j \frac{2\pi d}{z_k} \frac{B}{2} u} \left(H^{\dagger}(u, v) \vec{I}(u, v) \right)_k \tag{6.5}$$

$$S_{-B/2}(u, v) = \sum_{k=1}^{N} e^{+j \frac{2\pi d}{z_k} \frac{B}{2} u} \left(H^{\dagger}(u, v) \vec{I}(u, v) \right)_k \tag{6.6}$$

and finally:

$$i_L(x,y) = \mathcal{F}^{-1}\{S_{B/2}(u,v)\} \tag{6.7}$$

$$i_R(x,y) = \mathcal{F}^{-1}\{S_{-B/2}(u,v)\} \tag{6.8}$$

where i_L and i_R are the images seen by the left and right eye, respectively.

6.4.2 Displaying the stereoscopic pair
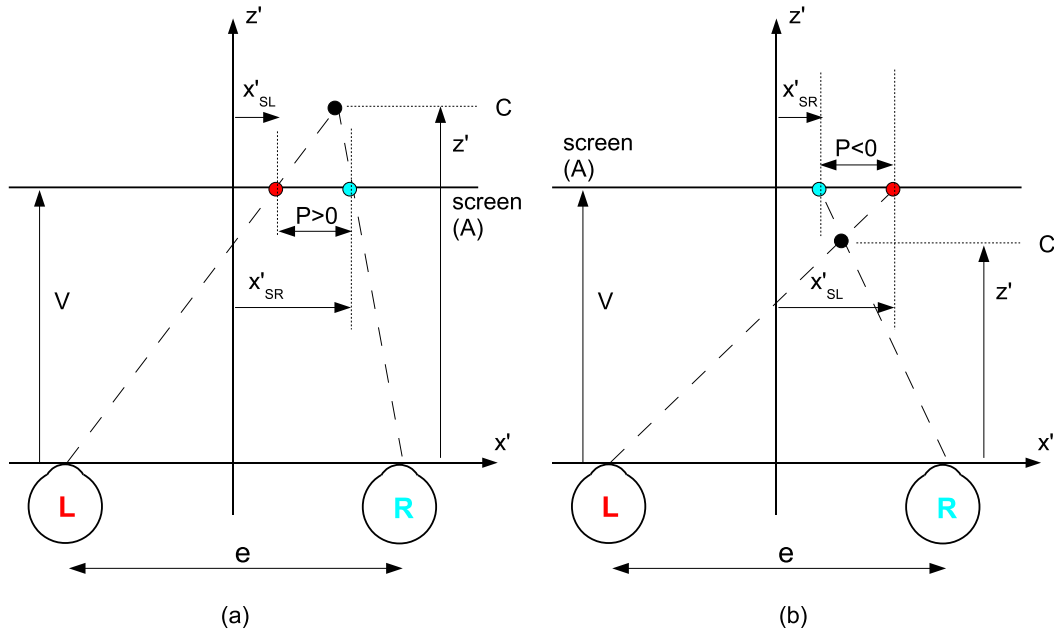
Figure 6.7: Perceived depth and parallax. V is the viewing distance (distance between the viewer and the screen), e is the inter-ocular distance and $P = X'_{SR} - X'_{SL}$ is the horizontal parallax. For the configuration shown in (a), parallax is positive, accomodation (A) occurs at distance V (viewing screen) while convergence (C) occurs behind the screen $(z' > V)$. For the configuration shown in (b), parallax is negative, accomodation (A) occurs at distance V (viewing screen) while convergence (C) occurs in-front of the screen $(z' < V)$.

When displaying the stereoscopic pair (see Figure 6.7), the parallax in the screen, $P = X'_{SR} - X'_{SL}$, is related with the disparity between the captured, in this case reconstructed, right and left images, $X_{CR} - X_{CL}$ [Broberg, 2011]. Here, X'_{SR} and X'_{SL} are the x-coordinates of the image of the object as perceived, respectively, by the right and left eye of an observer when displayed on the screen, while X_{CR} and X_{CL} correspond to the x-coordinates of the image of the object as captured by the stereo-camera.

In our approach the synthesis of a stereo pair of images is equivalent to the capture of the scene by a stereo camera in *parallel configuration* (i.e., the optical axes of both cameras are parallel). Under this parallel configuration, it can be seen that the two views always produce a negative horizontal disparity between the captured images (Figure 6.7(b))[1] (see below). Note that zero disparity in a captured 3D scene, under a *parallel camera configuration*, corresponds to a point at infinite distance.

Let M be the magnification between acquisition and displaying, i.e.,

$$M = \frac{screen\,width}{sensor\,width}, \tag{6.9}$$

then the parallax in the viewing screen will be related to the disparity between the acquired images [Smith and Collar, 2012] by:

$$P = X'_{SR} - X'_{SL} = M\left(X_{CR} - X_{CL}\right) \tag{6.10}$$

In our case, this leads to a negative parallax in the screen:

$$P = X'_{SR} - X'_{SL} = M\left(X_{CR} - X_{CL}\right) = M\left(\frac{-b_x d}{z} - \frac{b_x d}{z}\right)$$
$$= -M\left(\frac{2b_x d}{z}\right) = -M\frac{Bd}{z} \tag{6.11}$$

This way, when the stereo pair is displayed on a screen, the perceived depth can be deduced considering similar triangles in Figure 6.7,

$$\frac{P}{e} = \frac{z' - V}{z'} \tag{6.12}$$

then the perceived depth according to parallax is:

$$z' = \frac{e}{e - P}V. \tag{6.13}$$

From Equation (6.11) and Equation (6.13), the perceived depth is then given by:

$$z' = \frac{e}{e + M\frac{Bd}{z}}V. \tag{6.14}$$

When considering *parallel camera configuration* for acquisition, the infinity in real world ($z = \infty$) is placed at the screen level ($z' = V$) and everything else is perceived in front of the screen ($P < 0$, see Figure 6.7(b)). The objects appear to float between the observer and the screen plane (*pop-out image*).

---

[1] Positive parallax (Figure 6.7(a)) can only be attained when capturing the scene under a *toed-in* camera configuration

It is important to avoid excessive negative parallax because it causes excessive eye-convergence in-front of the screen, too close to the viewer, and can cause discomfort. The amount of parallax can be adjusted by selecting the value of B through an appropriate value of b_x.

It is worth noticing, although we are not going to implement it here, that it is possible to set convergence after acquisition, i.e., change the camera configuration from parallel to toed-in. The conventional method to do this is known as *horizontal image translation* (HIT) and consists in digitally applying an equal horizontal shift to each of the left/right views. This way the position of a particular object in the 3D scene can be perceived at the screen plane or even behind it [Broberg, 2011, Smith and Collar, 2012].

6.4.3 Results

Once the stereoscopic pair is generated, the left and right images can be displayed in different ways: cross-eyed viewing method [1], anaglyphs, parallax barrier displays and polarized displays among others [Iizuka, 2013]. We present here the stereoscopic pairs (right and left images) along with red/cyan color anaglyphs implemented in MATLAB[2], from the left and right reconstructed views.

To begin with, we present a synthetic multi-focus image stack from *peppers* in Figure 6.8. Then, the right and left stereo-pair is synthesized as described in the previous section, considering $b_x=-1mm$ and $b_x=1mm$ respectively. In Figure 6.9 the cross-eyed stereo pair is presented. With some practice, the fused image is perceived by deliberately crossing one's eyes until the two images come together. The same pair is shown in Figure 6.10 as an anaglyph.

For real scenes multi-focus stacks, we acquired the *dollar* set shown in Figure 6.11 using the optical system in Chapter 2. The reconstructed perspective shifts to the right ($b_x=-1.2mm$)and to the left ($b_x=1.2mm$) are shown in Figure 6.12. The corresponding anaglyph of *dollar* is shown in Figure 6.13. We also considered a more continuous real scene by means of *cat* set (see *Visualization acquired cat stack*). In this last case, we reconstructed the right and left images for the values ($b_x=-0.4mm$)

---

[1]The cross-eyed viewing method can be used on stereoscopic pictures where the left image is on the right side and vice versa (recall Figure 6.5(c)). Then crossing the eyes a third image should appear in the middle where left and right images merge together into a 3D image.

[2]The procedure is very simple, the two images of a stereo pair are combined into a single image by coding them in two different colors: the left image in red, and the right image in cyan. To accomplish this, we made the green and blue channel of the right image zero (resulting a red channel image only) and the red channel of the left image set to zero (resulting a blue and green channels image only). Alternatively, the anaglyphs as well as images for shutter glasses or parallax barrier displays can be generated from the left and right images by means of a freeware software like StereoPhoto Maker (http://stereo.jpn.org/eng/stphmkr/index.html).

and (b_x=0.4mm) respectively (see Figure 6.14). Finally, the red/cyan anaglyph is depicted in Figure 6.15.
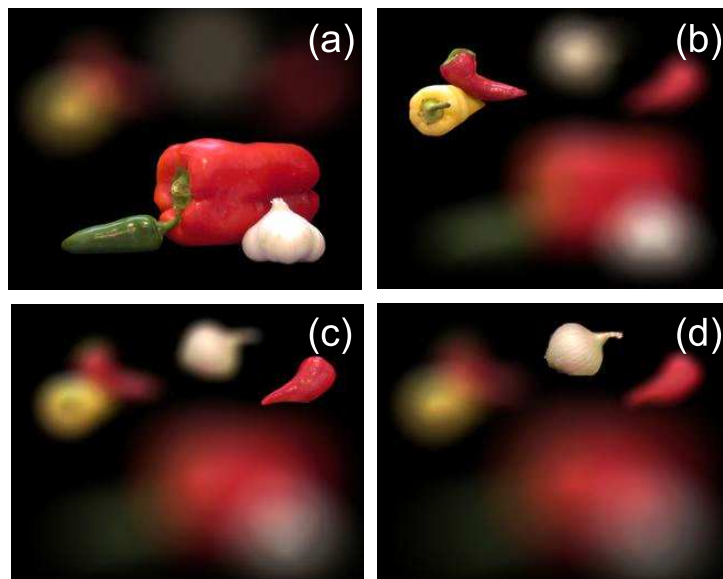
Figure 6.8: Synthetic defocused images generated from *"peppers"* (same stack as Figure 5.8). System focused at (a) $z_1 = 60mm$, (b) $z_2 = 90mm$,(c) $z_3 = 120mm$, and (d) $z_4 = 150mm$.
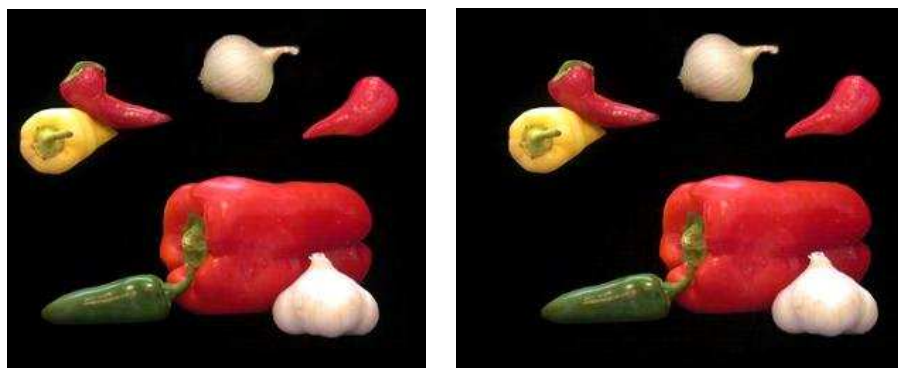
Figure 6.9: Synthetic stereoscopic pair for *peppers* set. Right and left images for cross-eyed view.
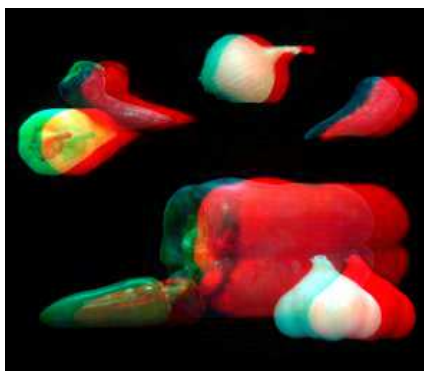
Figure 6.10: Color anaglyph for *peppers* set.(Red/cyan glasses are needed.)
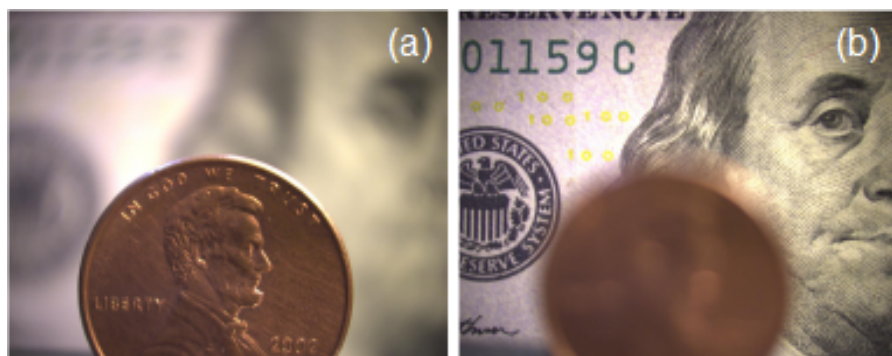
Figure 6.11: Acquired multi-focus image stack *dollar*. System focused at (a) $z_1 = 104mm$, (b) $z_2 = 148mm$.

Figure 6.12: Stereoscopic pair from *dollar* set. Right and left images for cross-eyed view.

Figure 6.13: Color anaglyph for *dollar* set. (Red/cyan glasses are needed.)

Figure 6.14: Stereoscopic pair from *cat* set. Right and left images for cross-eyed view.

Figure 6.15: Red cyan analgyph from *cat* set.(Red/cyan glasses are needed.)

Chapter 7

Conclusions

In many situations, the optical imaging system limits the depth-of-field of a scene, which makes it impossible to acquire the three-dimensional scene entirely in focus from a single-shot image. A possible solution consists of acquiring a multi-focus image stack and then computationally reconstruct the *all-in-focus* image (*extended depth-of-field*).

In the present work, we have implemented an optical system with an *electrically focus-tunable lens* (ETL) for the acquisition of multi-focus stacks which does not require image registration pre-processing. This way, the information of a three-dimensional scene is acquired in a time-sequential capture (the multi-focus stack) of high spatial resolution images (camera sensor resolution). Even though the temporally multiplexed approach comes at the cost of losing the instant capture of the scene, the development of new generations of ETLs with increasingly fast responses [Lu and Hua, 2015] that allow high-speed z-axis scanning of the scene, is bridging the gap between high spatial resolution and instant capture.

We developed post-capture image processing algorithms based on a linear image formation model that, besides the all-in-focus image reconstruction, allows for the synthesis of images with other novel features of interest. Operating in Fourier Domain, it is possible to handle optical settings such as pupil reshaping and viewpoint changes. The possibility of synthesizing different viewpoints of the scene, naturally led to the synthesis of stereoscopic pairs of images. In order to test the proposed algorithms performance, we implemented both qualitative and quantitative comparisons with state-of-the-art methods showing that our method outperforms others for severely defocused image stacks.

Following this Chapter, we list the publications and conference papers directly related to this work. We are currently working on an expanded scientific article for Optical Engineering, where we extent the content in [Alonso, 2016b], that was

recently presented at the conference *SPIE Optics + Photonics* about the synthesis of stereoscopic pairs of images from a multi-focus stack.

7.1 Future lines of research

Interesting lines of work can emerge from the possibility of synthesizing images with a different pupil from a different perspective at the same time. Additionally, the reconstruction of a series of perspective shifts (up, down, left, right, etc) to make an array of perspectives and the combination of this approach with that of integral imaging [Xiao et al., 2013] might allow to implement, with a proper lenslet array, an autostereoscopic display.

Other future line of research is in the field of microscopy. In conventional microscopy setups, differently focused images are obtained by mechanically changing the distance between the specimen and the objective lens to refocus at different depths. This brings about two disadvantages: first, neither moving the specimen nor moving the objective is convenient or fast enough. Secondly, the most important shortcoming for some applications in biology is that the motion may significantly disturb the specimen [Botcherby et al., 2007, Wu et al., 2012, Abrahamsson et al., 2013].

An alternative focusing method that does not require mechanical displacement of the specimen is clearly preferable [Botcherby et al., 2007]. In this regard, some approaches to include electrically focus-tunable lenses (ETLs) have been made with promising results [Martínez-Corral et al., 2015a, Martínez-Corral et al., 2015b]. This leads us to extent the proposed method to implement it in microscopy optical systems.

In particular, modern biological microscopy is also slowly moving away from imaging small samples, flattened between a slide and coverslip, toward three-dimensional cell cultures, to study development and physiology in more natural environments. With their low cost, simple construction and control, and wide focus-tuning ranges, focus-tunable lenses are especially well suited for microscopy applications (`http://www.photonics.com/Article.aspx?AID=57323`). In this regard, the extension of the method towards its potential application in three-dimensional microscopy for imaging thick biological specimens constitutes a challenging future line of research. Some contacts with researchers from *Facultad de Ciencias* and *Instituto de Investigaciones Biológicas Clemente Estable* (IIBCE) have been made in this direction in order to adapt our method for its implementation in the study of *multicellular tumor spheroids* (MCTS). Unlike classical monolayer-based models, MCTS strikingly mirror the 3D cellular context and therapeutically relevant pathophysiological gradients

of in vivo tumors, which makes them a promising anti-cancer therapy test [Carver et al., 2014, Hirschhaeuser et al., 2010].

Publications

List of publications and presentations directly related to this thesis:

Journal articles

- Alonso, J. R., Fernández, A., Ayubi, G. A., and Ferrari, J. A. (2015). All-in-focus image reconstruction under severe defocus. *Optics Letters*, 40(8):1671–1674

- Alonso, J. R., Fernández, A., and Ferrari, J. A. (2016). Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack. *Applied Optics*, 55(9):2380–2386

Conference papers

- Alonso, J. and Ferrari, J. (2015). From frequency domain multi-focus fusion to focus slicing. In *Frontiers in Optics*, pages FTh3G–5. Optical Society of America, OSA

- Alonso, J. (2016a). Fourier domain method for extended depth-of-field from a multi-focus image stack. In *Imaging Systems and Applications*, pages IW3F–3. Optical Society of America

- Alonso, J. (2016b). Stereoscopic 3D scene synthesis from a monocular camera with an electrically tunable lens. In *SPIE Optics + Photonics. Optical Engineering + Applications, Optics and Photonics for Information Processing X*, pages 9970–18. Society of Photo-Optical Instrumentation Engineers, SPIE

- Alonso, J. (2016c). Synthetically reshaped aperture for postacquisition three-dimensional scene refocusing from a multi-focus image stack. In *Computational Optical Sensing and Imaging*, pages JT3A–48. Optical Society of America

Appendix A

Moore-Penrose pseudoinverse solution to the inverse problem.

This appendix is devoted to show that the minimum norm solution provided by the Moore-Penrose pseudoinverse gives an exact solution to Equation (3.30) when dealing with the case $(u, v) = (0, 0)$ for which H matrix is non-invertible.

Let us start from the image formation model in the Fourier domain:

$$\vec{I}(u, v) = H(u, v)\vec{F}(u, v). \tag{A.1}$$

The Moore-Penrose pseudo-inverse $H^\dagger$ [Ben-Israel and Greville, 2003, Barata and Hussein, 2012, Favaro and Soatto, 2007] is characterized by the following properties:

(i) $HH^\dagger H = H$

(ii) $H^\dagger H H^\dagger = H^\dagger$

(iii) $\left(HH^\dagger\right)^* = HH^\dagger$

(iv) $\left(H^\dagger H\right)^* = H^\dagger H$

and provides the set of vectors that minimize the Euclidean norm $\left\| H(u, v)\vec{F}(u, v) - \vec{I}(u, v) \right\|$ in the least squares sense:

$$\vec{F}^{MP}(u, v) = H^\dagger(u, v)\vec{I}(u, v) + \left(\mathbb{1}_N - H^\dagger(u, v)H(u, v)\right)\vec{w}, \tag{A.2}$$

with arbitrary vector $\vec{w}$ and $N \times N$ identity matrix $\mathbb{1}_N$. When $H(u, v)$ is invertible then $H^\dagger(u, v) = H^{-1}(u, v)$ and Equation (A.2) reduces to $\vec{F}(u, v) = H^{-1}(u, v)\vec{I}(u, v)$. For a given frequency pair (u, v) the first term in Equation (A.2) is the minimal norm vector within the minimal set and belongs to $ran(H^\dagger(u, v))$, i.e., *range* [Horn and

Johnson, 2012] of $H^\dagger(u,v)$. By using the property in Item (i) it is clear that the result of applying $H(u,v)$ by the left over the second term in Equation (A.2) is the null vector regardless of $\vec{w}$, so $\left(\mathbb{1}_N - H^\dagger(u,v)H(u,v)\right)\vec{w}$ belongs to $ker(H(u,v))$, i.e., the *kernel* [Horn and Johnson, 2012] of $H(u,v)$. Since $ran(H^\dagger(u,v)) = ker(H(u,v))^\perp$ it results that the terms in the right hand side (r.h.s.) of Equation (A.2) are mutually orthogonal. This orthogonality is a crucial point when considering the possible combinations of the components of $\vec{F}(u,v)$ allowed through our method.

A.1 DC component of H for arbitrary N

A non-invertible case of $H(u,v)$ of great interest arises for the DC component $(u,v) = (0,0)$. Since the components of matrix H are OTFs, $H_{kk'}(0,0) = 1$ so $H(0,0)$ takes the form of an $N \times N$ all-ones matrix:

$$H(0,0) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} = \mathbb{J}_N, \tag{A.3}$$

where we denote by $\mathbb{J}_N$ the $N \times N$ all-ones matrix which for the cases of of interest $(N \geq 2)$ is non-invertible. Some properties of $\mathbb{J}_N$ that will result useful are [Horn and Johnson, 2012]:

(a) The eigenvalues of $\mathbb{J}_N$ are: N with multiplicity 1 and 0 with multiplicity $N-1$.

(b) The eigenvector associated with N is the all-ones vector $\vec{e}$: $\mathbb{J}_N\vec{e} = N\vec{e}$.

(c) $(\mathbb{J}_N)^k = N^{k-1}\mathbb{J}_N$

By means of Equation (A.3), Equation (A.2) then takes the form:

$$\vec{F}^{MP}(0,0) = \mathbb{J}_N^\dagger \vec{I}(0,0) + \left(\mathbb{1}_N - \mathbb{J}_N^\dagger \mathbb{J}_N\right)\vec{w}. \tag{A.4}$$

Due to normalization of radiant energy between images of the stack we have (recall Equation (A.5)):

$$\vec{I}(0,0) = \frac{1}{N}\left(\sum_k I_k(0,0)\right)\vec{e}, \tag{A.5}$$

and Equation (A.4) simplifies to:

$$\vec{F}^{MP}(0,0) = \frac{1}{N}\left(\sum_k I_k(0,0)\right)\mathbb{J}_N^\dagger\vec{e} + \left(\mathbb{1}_N - \mathbb{J}_N^\dagger\mathbb{J}_N\right)\vec{w}. \tag{A.6}$$

It can be demonstrated by direct substitution in the defining properties Items (i) to (iv) and by applying Item (c) that $\mathbb{J}_N^\dagger$ can be expressed as:

$$\mathbb{J}_N^\dagger = \frac{1}{N^2}\mathbb{J}_N. \tag{A.7}$$

Equation (A.6) can then be further simplified to:

$$\vec{F}^{MP}(0,0) = \frac{1}{N}\left(\sum_k I_k(0,0)\right)\frac{1}{N^2}\mathbb{J}_N\vec{e} + \left(\mathbb{1}_N - \mathbb{J}_N^\dagger\mathbb{J}_N\right)\vec{w}, \tag{A.8}$$

and by using Item (b) we obtain:

$$\vec{F}^{MP}(0,0) = \frac{1}{N}\left(\sum_k I_k(0,0)\right)\frac{1}{N}\vec{e} + \left(\mathbb{1}_N - \mathbb{J}_N^\dagger\mathbb{J}_N\right)\vec{w}. \tag{A.9}$$

The first term on the r.h.s. of Equation (A.9) is then proportional to the all-ones vector $\vec{e}$. If we consider combinations of the components of $\vec{F}^{MP}$ whose DC component can be written as $\vec{e}\cdot\vec{F}^{MP}(0,0)$ we obtain:

$$\vec{e}\cdot\vec{F}^{MP}(0,0) = \vec{e}\cdot\frac{1}{N}\left(\sum_k I_k(0,0)\right)\frac{1}{N}\vec{e} + \vec{e}\cdot\left(\mathbb{1}_N - \mathbb{J}_N^\dagger\mathbb{J}_N\right)\vec{w} = \frac{1}{N}\left(\sum_k I_k(0,0)\right), \tag{A.10}$$

where as the first term in Equation (A.9) results collinear with $\vec{e}$ and we had that the terms in the r.h.s. of Equation (A.2) are mutually orthogonal, then the second term in Equation (A.10) vanishes by virtue of the orthogonality of the second r.h.s. term in Equation (A.9) with $\vec{e}$. In this regard, as long as we consider combinations like Equation (3.30) or Equation (4.6), which for its DC component behaves like $\vec{e}\cdot\vec{F}^{MP}(0,0)$ (e.g. all-in-focus image reconstruction), the arbitrary term in $\vec{w}$ found in Equation (A.2) will play no role when projected over $\vec{e}$ and the approximation of $\vec{e}\cdot\vec{F}^{MP}(0,0)$ by $\vec{e}\cdot H^\dagger(0,0)\vec{I}(0,0)$ is exact.

A.2 Graphical representation for the case $N = 2$

In those cases where we only deal with a stack of $N = 2$ images, i.e., foreground and background, it is possible to graphically represent the previous discussion following

the representation depicted in [Schechner et al., 2000]. Let us explicitly write the image formation model (A.1) for $N = 2$ images:

$$\left(\begin{array}{c} I_1(u,v) \\ I_2(u,v) \end{array} \right) = \left(\begin{array}{cc} 1 & H_{12}(u,v) \\ H_{12}(u,v) & 1 \end{array} \right) \left(\begin{array}{c} F_1(u,v) \\ F_2(u,v) \end{array} \right), \qquad \text{(A.11)}$$

that is:

$$I_1(u,v) = F_1(u,v) + H_{12}(u,v)F_2(u,v) \qquad \text{(A.12a)}$$

$$I_2(u,v) = H_{12}(u,v)F_1(u,v) + F_2(u,v). \qquad \text{(A.12b)}$$

For center-symmetric pupils as the circular pupil in our case, H_{12} is real and taking the real part of Equations (A.12a) and (A.12b) we obtain:

$$Re\{I_1(u,v)\} = Re\{F_1(u,v)\} + H_{12}(u,v)Re\{F_2(u,v)\} \qquad \text{(A.13a)}$$

$$Re\{I_2(u,v)\} = H_{12}(u,v)Re\{F_1(u,v)\} + Re\{F_2(u,v)\}, \qquad \text{(A.13b)}$$

and analogously for the imaginary part of the equation system. Each one of the Equations (A.13a) and (A.13b) can be plotted as a line in the $\{Re\{F_1(u,v)\}, Re\{F_2(u,v)\}\}$ space (Figure A.1) which will be denoted $\{F_1(u,v), F_2(u,v)\}$. The solution of the system corresponds to the intersection point of the lines (Figure A.1(a)) which is unique while $H_{12}(u,v) \neq 1$. The case $H_{12}(u,v) = 1$ corresponds only to $(u,v) = (0,0)$, for which the determinant of matrix H in Equation (A.11):

$$|H(0,0)| = 1 - H_{12}^2(0,0) \qquad \text{(A.14)}$$

vanishes and H is no longer invertible. Under these circumstances, the picture in $\{F_1(0,0), F_2(0,0)\}$ plane is of coinciding lines representing Equations (A.13a) and (A.13b) (Figure A.1(b)) and giving infinite solutions to the system. The line containing any of the possible solutions passes through the points $(I_1(0,0), 0)$ and $(0, I_2(0,0))$; according to Equation (A.5) $I_1(0,0) = I_2(0,0) = \frac{1}{N}\left(\sum_k I_k(0,0)\right)$ so the normal to the line is given by $\vec{e} = (1,1)$. According to the normal form of a line, the possible solutions $\vec{F}^{MP}(0,0)$ satisfy:

$$\vec{e} \cdot \vec{F}^{MP}(0,0) = C \qquad \text{(A.15)}$$

where C is a constant,that is, any of the possible solutions gives the same projection over $\vec{e}$ (C is given by $\vec{e} \cdot \mathbb{J}_2^\dagger \vec{I}(0,0)$).
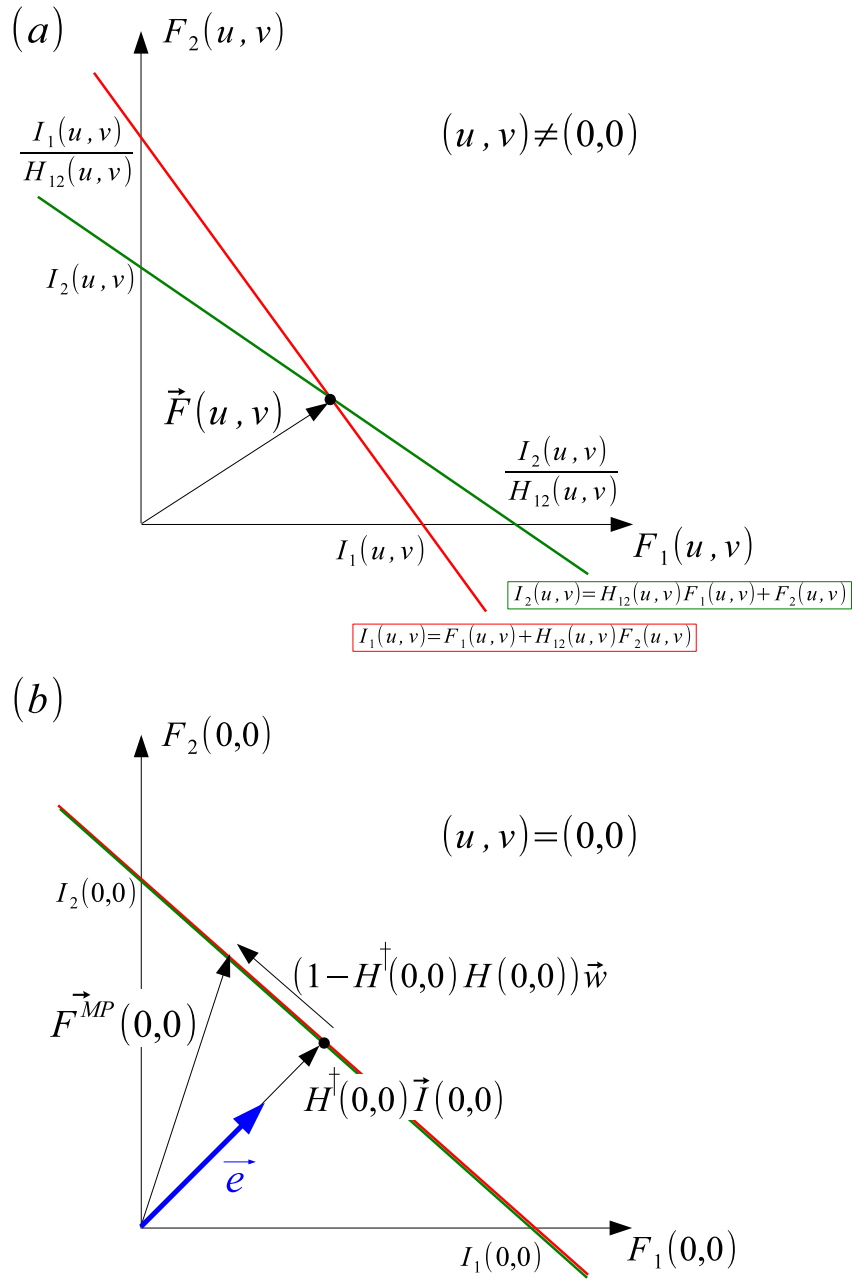
Figure A.1: Graphical solution for the system of eqs. (A.13a) and (A.13b). (a) $(u, v) \neq (0, 0)$ (b) $(u, v) = (0, 0)$.

Appendix B

From Frequency Domain Multi-Focus Fusion to Focus Slicing

In the present appendix we propose a method [Alonso and Ferrari, 2015] to estimate the focus slices (i.e. the in-focus regions of each of the acquired images of a multi-focus stack), which takes advantage of the all-in-focus image reconstruction scheme developed in Chapter 3.

The set of images over which we will estimate the focus slices is shown in Figure B.1(a-c). Each image will contain both in-focus information of the scene as well as out-of-focus contributions.

Let us start by recalling the image formation model in Equation (3.12) where the i_k image taken with the system focused at axial distance $z = z_k$ can be described by the following equation:

$$i_k(x,y) = f_k(x,y) + \sum_{k' \neq k} h_{kk'}(x,y) * f_{k'}(x,y), \qquad \text{(B.1)}$$

Figure B.1: $(a - c)$ acquired stack of $N = 3$ real images focused at 111mm, 119mm and 135mm, respectively. (d) Fused image (all-in-focus).

where f_k is the in-focus region of i_k (i.e. f_k is the focus slice corresponding to z_k). The fused image $s(x, y)$ corresponds to the sum of the focus sl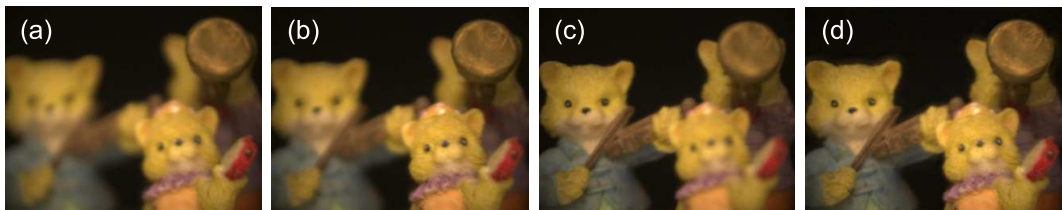ices (in-focus components of the different images), $s(x, y) = \sum_{k=1}^{N} f_k(x, y)$, and it is shown in Figure B.1(d).

Recall now that the reconstruction of the all-in-focus image in frequency domain is given by Equation (3.31)

$$S(u, v) = \sum_{k=1}^{N} F_k(u, v) = \vec{e} \cdot \vec{F}(u, v) = \vec{e} \cdot \left(H^\dagger(u, v) \vec{I}(u, v) \right), \qquad \text{(B.2)}$$

where $\vec{e}$ is the vector of length N whose elements are ones and $H^\dagger$ is the Moore-Penrose pseudo-inverse, which reduces to H^{-1} when H is invertible. The all-in-focus image $s(x, y)$ is obtained through inverse Fourier transforming Equation (B.2) to the spatial domain (Equation (3.33))

$$s(x, y) = \mathcal{F}^{-1}\{S(u, v)\}. \qquad \text{(B.3)}$$

For the DC component (i.e. $(u, v) = (0, 0)$), $H(0, 0)$ is a singular matrix, and the only constraint on the focus slices is $\sum_{k=1}^{N} F_k(0, 0) = S(0, 0)$, so the retrieval of the DC components of the focus slices is a (N-1)-fold degenerate problem. Then, it is not possible to obtain the focus slices f_k considering only the frequency domain information. This degeneracy was already stated in Item (a) in Appendix A.

As a first approximation to each focus slice in Fourier domain we can use the minimum norm vector from the Moore-Penrose (MP) pseudo-inverse solution (first term in Equation (A.2))

$$F_k^{MP} = \left(H^\dagger \vec{I} \right)_k, \qquad \text{(B.4)}$$

which in particular, assigns the value $\dfrac{S(0, 0)}{N}$ to each $F_k(0, 0)$:

$$F_k^{MP}(0, 0) = \frac{S(0, 0)}{N}. \qquad \text{(B.5)}$$

The proposed method to estimate the focus slices takes as a starting point the Moore-Penrose Fourier focus slices from Equation (B.4) and is described in the following.

Let s^{HP} be the result of high-pass filtering the all-in-focus image

$$s^{HP} = \mathcal{F}^{-1}\{S(1 - G)\} \qquad \text{(B.6)}$$

where G is a low-pass filter (note that in the absence of the low-pass filter, i.e. $G = 0$, the usual all-in-focus image is retrieved from Equation (B.6)). Assuming that the high frequency content of the focus slices was correctly retrieved from Equation (B.4)

we can define the high-pass filtered focus slices f_k^{HP} as

$$f_k^{HP} = \mathcal{F}^{-1}\{F_k^{MP}(1-G)\}. \tag{B.7}$$

Next we make a block-based pixel by pixel comparison between s^{HP} and f_k^{HP}, resulting in a mask for each k:

$$m_k(i,j) = \begin{cases} 1 & \text{if } k = \arg\min_{k'}\left\langle\left|f_{k'}^{HP}(i,j) - s^{HP}(i,j)\right|^2\right\rangle_{B\times B} \\ \\ 0 & \text{otherwise} \end{cases} . \tag{B.8}$$

In the last expression, $|\cdot|^2$ is the square of the distance in RGB space between s^{HP} and f_k^{HP}; $\langle\cdot\rangle_{B\times B}$ is the mean over a square block of side B; and $\arg\min_{k'}$ selects the index k for which the enclosed expression achieves its minimum.

Let us now introduce ω_k, the result of applying each of the m_k masks to the all-in-focus image:

$$\omega_k = m_k s. \tag{B.9}$$

Each of these new images ω_k corresponds to a segmentation of the all-in-focus image. We now retrieve the low frequency information of each ω_k and finally the estimated (new) focus slices transforms can be reconstructed as

$$F^{new} = \Omega_k G + F_k^{MP}(1-G), \tag{B.10}$$

where $\Omega_k = \mathcal{F}\{\omega_k\}$. As a test of consistency note that new focus slices give rise to the same fusion image: from Equation (B.9) and the definition in Equation (B.8) it is clear that

$$\sum_{k=1}^N \omega_k = s, \tag{B.11}$$

which in Fourier domain reads:

$$\sum_{k=1}^N \Omega_k = S; \tag{B.12}$$

note also from Equation (B.4) and Equation (B.2)

$$\sum_{k=1}^N F_k^{MP} = S, \tag{B.13}$$
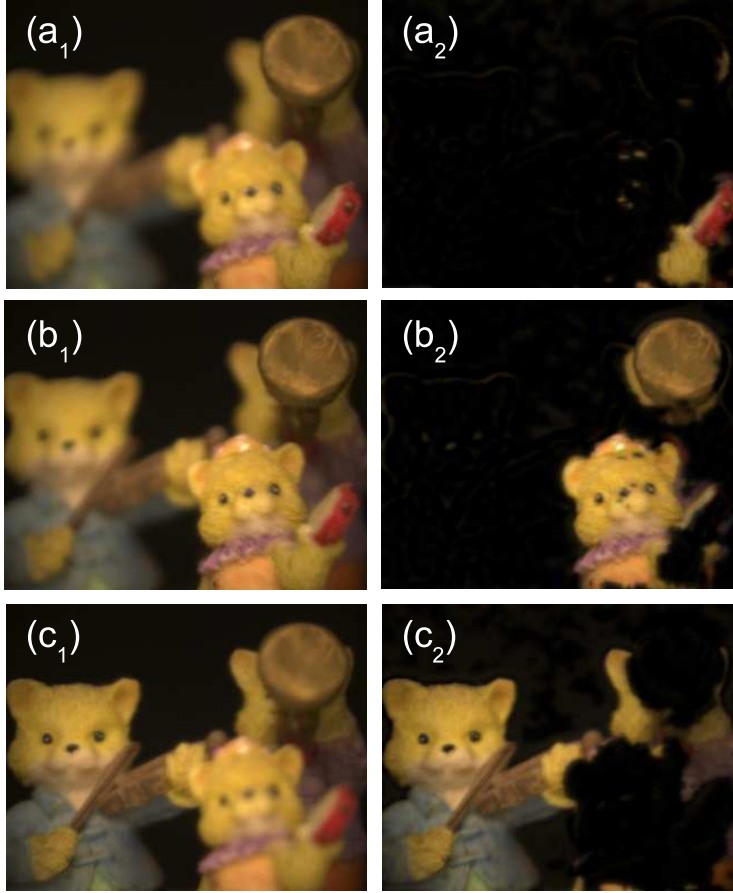
Figure B.2: Focus slices. $(a_1 - c_1)$ acquired stack of $N = 3$ real images focused at 111mm, 119mm and 135mm, respectively. $(a_2 - c_2)$ Estimated focus slices of the stack. We considered a 11×11px block for the computation of the mask and a first order Butterworth low-pass filter with radius $\sigma = 7N$.

so adding the contributions in Equation (B.10) from every slice we have:

$$\sum_{k=1}^{N} F_k^{new} = \left(\sum_{k=1}^{N} \Omega_k \right) G + \left(\sum_{k=1}^{N} F_k^{MP} \right) (1 - G) = SG + S(1 - G) = S, \quad \text{(B.14)}$$

recovering the same all-in-focus image from the beginning.

Finally, the space domain focus slices f_k^{new} are obtained through inverse Fourier transforming Equation (B.10):

$$f_k^{new} = \mathcal{F}^{-1}\{\Omega_k G + F_k^{MP}(1 - G)\}. \quad \text{(B.15)}$$

The result of applying the previous method over the image stack is shown in Figure B.2.

112

Bibliography

[Abrahamsson et al., 2013] Abrahamsson, S., Chen, J., Hajj, B., Stallinga, S., Katsov, A. Y., Wisniewski, J., Mizuguchi, G., Soule, P., Mueller, F., Darzacq, C. D., et al. (2013). Fast multicolor 3d imaging using aberration-corrected multi-focus microscopy. *Nature Methods*, 10(1):60–63.

[Allen and Triantaphillidou, 2012] Allen, E. and Triantaphillidou, S. (2012). *The Manual of Photography and Digital Imaging*. CRC Press.

[Alonso, 2016a] Alonso, J. (2016a). Fourier domain method for extended depth-of-field from a multi-focus image stack. In *Imaging Systems and Applications*, pages IW3F–3. Optical Society of America.

[Alonso, 2016b] Alonso, J. (2016b). Stereoscopic 3D scene synthesis from a monocular camera with an electrically tunable lens. In *SPIE Optics + Photonics. Optical Engineering + Applications, Optics and Photonics for Information Processing X*, pages 9970–18. Society of Photo-Optical Instrumentation Engineers, SPIE.

[Alonso, 2016c] Alonso, J. (2016c). Synthetically reshaped aperture for postacquisition three-dimensional scene refocusing from a multi-focus image stack. In *Computational Optical Sensing and Imaging*, pages JT3A–48. Optical Society of America.

[Alonso and Ferrari, 2015] Alonso, J. and Ferrari, J. (2015). From frequency domain multi-focus fusion to focus slicing. In *Frontiers in Optics*, pages FTh3G–5. Optical Society of America, OSA.

[Alonso et al., 2015] Alonso, J. R., Fernández, A., Ayubi, G. A., and Ferrari, J. A. (2015). All-in-focus image reconstruction under severe defocus. *Optics Letters*, 40(8):1671–1674.

[Alonso et al., 2016] Alonso, J. R., Fernández, A., and Ferrari, J. A. (2016). Reconstruction of perspective shifts and refocusing of a three-dimensional scene from a multi-focus image stack. *Applied Optics*, 55(9):2380–2386.

[Aslantas and Pham, 2007] Aslantas, V. and Pham, D. (2007). Depth from automatic defocusing. *Optics Express*, 15(3):1011–1023.

[Barata and Hussein, 2012] Barata, J. C. A. and Hussein, M. S. (2012). The Moore–Penrose pseudoinverse: A tutorial review of the theory. *Brazilian Journal of Physics*, 42(1-2):146–165.

[Ben-Israel and Greville, 2003] Ben-Israel, A. and Greville, T. N. (2003). *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media.

[Bilgen et al., 2013] Bilgen, K., Üstün, M., Karakahya, M., Isik, S., Sengül, S., Çetinkünar, S., and Küçükpinar, T. H. (2013). Comparison of 3d imaging and 2d imaging for performance time of laparoscopic cholecystectomy. *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, 23(2):180–183.

[Bishop and Favaro, 2012] Bishop, T. E. and Favaro, P. (2012). The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):972–986.

[Born and Wolf, 2000] Born, M. and Wolf, E. (2000). *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. CUP Archive.

[Botcherby et al., 2007] Botcherby, E. J., Juskaitis, R., Booth, M. J., and Wilson, T. (2007). Aberration-free optical refocusing in high numerical aperture microscopy. *Optics Letters*, 32(14).

[Broberg, 2011] Broberg, D. K. (2011). Guidance for Horizontal Image Translation (HIT) of high definition stereoscopic video production. In *IS&T/SPIE Electronic Imaging*, pages 78632F–78632F. International Society for Optics and Photonics.

[Burt and Adelson, 1983] Burt, P. and Adelson, E. (1983). The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540.

[Carver et al., 2014] Carver, K., Ming, X., and Juliano, R. L. (2014). Multicellular tumor spheroids as a model for assessing delivery of oligonucleotides in three dimensions. *Molecular Therapy. Nucleic Acids*, 3(3):e153.

[Cat, 2013] Cat, J. (2013). *Maxwell, Sutton, and the Birth of Color Photography: A Binocular Study*. Springer.

[Favaro and Soatto, 2003] Favaro, P. and Soatto, S. (2003). Seeing beyond occlusions (and other marvels of a finite lens aperture). In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–579. IEEE.

[Favaro and Soatto, 2007] Favaro, P. and Soatto, S. (2007). *3-D shape estimation and image restoration: Exploiting defocus and motion-blur.* Springer Science & Business Media.

[Fehn et al., 2006] Fehn, C., De La Barré, R., and Pastoor, S. (2006). Interactive 3-D TV-concepts and key technologies. *Proceedings of the IEEE*, 94(3):524–538.

[Ferdig et al., 2015] Ferdig, R., Blank, J., Kratcoski, A., and Clements, R. (2015). Using stereoscopy to teach complex biological concepts. *Advances in Physiology Education*, 39(3):205–208.

[Feruglio et al., 2013] Feruglio, P. F., Vinegoni, C., Fexon, L., Thurber, G., Sbarbati, A., and Weissleder, R. (2013). Noise suppressed, multifocus image fusion for enhanced intraoperative navigation. *Journal of Biophotonics*, 6(4):363–370.

[Fong et al., 2015] Fong, Y., Giulianotti, P. C., Lewis, J., Koerkamp, B. G., and Reiner, T. (2015). *Imaging and Visualization in The Modern Operating Room: A Comprehensive Guide for Physicians.* Springer.

[Funk, 2012] Funk, W. (2012). History of autostereoscopic cinema. In *IS&T/SPIE Electronic Imaging*, pages 82880R–82880R. International Society for Optics and Photonics.

[Ghosh et al., 2014] Ghosh, S., Schaefer, L., Schuster, D., and Preza, C. (2014). Further developments in addressing depth-variant 3d fluorescence microscopy imaging. In *SPIE BiOS*, pages 89490Q–89490Q. International Society for Optics and Photonics.

[Gonzalez et al., 2009] Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2009). *Digital Image Processing using MATLAB.* Gatesmark Publishing.

[Goodman, 1996] Goodman, J. W. (1996). *Introduction to Fourier Optics.* Roberts and Company Publishers.

[Greengard et al., 2006] Greengard, A., Schechner, Y. Y., and Piestun, R. (2006). Depth from diffracted rotation. *Optics Letters*, 31(2):181–183.

[Guo et al., 2012] Guo, L., Dai, M., and Zhu, M. (2012). Multifocus color image fusion based on quaternion curvelet transform. *Optics Express*, 20(17):18846–18860.

[Hadj and Blanc-Féraud, 2012] Hadj, S. B. and Blanc-Féraud, L. (2012). Modeling and removing depth variant blur in 3D fluorescence microscopy. In *2012 IEEE*

International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 689–692. IEEE.

[Harding, 2016] Harding, K. (2016). Multi-focus, high resolution inspection system for extended range applications. In *SPIE Commercial+ Scientific Sensing and Imaging*, pages 98680D–98680D. International Society for Optics and Photonics.

[Harman and De Regt, 1996] Harman, P. and De Regt, H. W. (1996). The Scientific Letters and Papers of James Clerk Maxwell, volume ii: 1862-1873.

[Held et al., 2012] Held, R. T., Cooper, E. A., and Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, 22(5):426–431.

[Hirschhaeuser et al., 2010] Hirschhaeuser, F., Menne, H., Dittfeld, C., West, J., Mueller-Klieser, W., and Kunz-Schughart, L. A. (2010). Multicellular tumor spheroids: an underestimated tool is catching up again. *Journal of Biotechnology*, 148(1):3–15.

[Horn and Johnson, 2012] Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press.

[Howard and Rogers, 2012] Howard, I. and Rogers, B. (2012). Perceiving in depth, volume 2: Stereoscopic Vision, no. 29.

[Howard and Rogers, 1995] Howard, I. P. and Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford University Press, USA.

[Huang et al., 2015] Huang, F.-C., Chen, K., and Wetzstein, G. (2015). The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Transactions on Graphics (TOG)*, 34(4):60.

[Huang and Jing, 2007] Huang, W. and Jing, Z. (2007). Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28(4):493–500.

[Iizuka, 2013] Iizuka, K. (2013). *Engineering Optics*, volume 35. Springer Science & Business Media.

[Jacobs et al., 2012] Jacobs, D. E., Baek, J., and Levoy, M. (2012). Focal stack compositing for depth of field control. *Stanford Computer Graphics Laboratory Technical Report*, 1:2012.

[Je and Park, 2013] Je, C. and Park, H.-M. (2013). Optimized hierarchical block matching for fast and accurate image registration. *Signal Processing: Image Communication*, 28(7):779–791.

[Jenkins and White, 1957] Jenkins, F. A. and White, H. E. (1957). *Fundamentals of optics*. Tata McGraw-Hill Education.

[Jesacher et al., 2015] Jesacher, A., Ritsch-Marte, M., and Piestun, R. (2015). Three-dimensional information from two-dimensional scans: a scanning microscope with postacquisition refocusing capability. *Optica*, 2(3):210–213.

[Kenneth, 1996] Kenneth, R. (1996). *Castleman, Digital Image Processing*. Prentice Hall Press, Upper Saddle River, NJ.

[Kim and Naemura, 2015] Kim, B. and Naemura, T. (2015). Blind depth-variant deconvolution of 3d data in wide-field fluorescence microscopy. *Scientific Reports*, 5.

[Kirshner et al., 2013] Kirshner, H., Aguet, F., Sage, D., and Unser, M. (2013). 3-d psf fitting for fluorescence microscopy: implementation and localization application. *Journal of microscopy*, 249(1):13–25.

[Kodama and Kubota, 2013] Kodama, K. and Kubota, A. (2013). Efficient reconstruction of all-in-focus images through shifted pinholes from multi-focus images for dense light field synthesis and rendering. *IEEE Transactions on Image Processing*, 22(11):4407–4421.

[Kubota and Aizawa, 2005] Kubota, A. and Aizawa, K. (2005). Reconstructing arbitrarily focused images from two differently focused images using linear filters. *ieee Transactions on Image Processing*, 14(11):1848–1859.

[Kubota et al., 2007] Kubota, A., Aizawa, K., and Chen, T. (2007). Reconstructing dense light field from array of multifocus images for novel view synthesis. *IEEE Transactions on Image Processing*, 16(1):269–279.

[Kubota et al., 1999] Kubota, A., Kodama, K., and Aizawa, K. (1999). Registration and blur estimation methods for multiple differently focused images. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 2, pages 447–451. IEEE.

[Kunert et al., 2013] Kunert, W., Storz, P., and Kirschniak, A. (2013). For 3D laparoscopy: a step toward advanced surgical navigation: how to get maximum benefit from 3D vision. *Surgical Endoscopy*, 27(2):696–699.

[Kuthirummal et al., 2011] Kuthirummal, S., Nagahara, H., Zhou, C., and Nayar, S. K. (2011). Flexible depth of field photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):58–71.

[Lam, 2011] Lam, E. Y. (2011). Computational photography: advances and challenges. In *SPIE Optical Engineering+ Applications*, pages 81220O–81220O. International Society for Optics and Photonics.

[Lam, 2015] Lam, E. Y. (2015). Computational photography with plenoptic camera and light field capture: tutorial. *JOSA A*, 32(11):2021–2032.

[Lambooij et al., 2009] Lambooij, M., Fortuin, M., Heynderickx, I., and IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):30201–1.

[Levin and Durand, 2010] Levin, A. and Durand, F. (2010). Linear view synthesis using a dimensionality gap light field prior. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1831–1838. IEEE.

[Levin et al., 2007] Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007). Image and depth from a conventional camera with a coded aperture. *ACM transactions on graphics (TOG)*, 26(3):70.

[Levoy, 2006] Levoy, M. (2006). Light fields and computational imaging. *IEEE Computer*, 39(8):46–55.

[Levoy, 2010] Levoy, M. (2010). Experimental platforms for computational photography (october, pg 81, 2010). *IEEE Computer Graphics and Applications*, 30(6):15–15.

[Li et al., 2011] Li, S., Yang, B., and Hu, J. (2011). Performance comparison of different multi-resolution transforms for image fusion. *Information Fusion*, 12(2):74–84.

[Liang et al., 2008] Liang, C.-K., Lin, T.-H., Wong, B.-Y., Liu, C., and Chen, H. H. (2008). Programmable aperture photography: multiplexed light field acquisition. In *ACM Transactions on Graphics (TOG)*, volume 27, page 55. ACM.

[Lin et al., 2013] Lin, X., Suo, J., Wetzstein, G., Dai, Q., and Raskar, R. (2013). Coded focal stack photography. In *Computational Photography (ICCP), 2013 IEEE International Conference on*, pages 1–9. IEEE.

[Lin et al., 2015] Lin, X., Wu, J., Zheng, G., and Dai, Q. (2015). Camera array based light field microscopy. *Biomedical Optics Express*, 6(9):3179–3189.

[Liu and Hua, 2011] Liu, S. and Hua, H. (2011). Extended depth-of-field microscopic imaging with a variable focus microscope objective. *Optics Express*, 19(1):353–362.

[Liu et al., 2015] Liu, Y., Liu, S., and Wang, Z. (2015). A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*, 24:147–164.

[Lu and Hua, 2015] Lu, S.-H. and Hua, H. (2015). Imaging properties of extended depth of field microscopy through single-shot focus scanning. *Optics Express*, 23(8):10714–10731.

[Mahajan and Díaz, 2016] Mahajan, V. N. and Díaz, J. A. (2016). Comparison of geometrical and diffraction imaging in the space and frequency domains. *Applied Optics*, 55(12):3241–3250.

[Mahmood and Choi, 2010] Mahmood, M. T. and Choi, T.-S. (2010). Focus measure based on the energy of high-frequency components in the S transform. *Optics Letters*, 35(8):1272–1274.

[Martínez-Corral et al., 2015a] Martínez-Corral, M., Doblas, A., Sánchez-Ortiga, E., Sola-Pikabea, J., and Saavedra, G. (2015a). Static real-time capture of 3D microscopy images. *Nature*, 1:R2.

[Martínez-Corral et al., 2014] Martínez-Corral, M., Dorado, A., Navarro, H., Llavador, A., Saavedra, G., and Javidi, B. (2014). From the plenoptic camera to the flat integral-imaging display. In *SPIE Sensing Technology+ Applications*, pages 91170H–91170H. International Society for Optics and Photonics.

[Martínez-Corral et al., 2015b] Martínez-Corral, M., Hsieh, P.-Y., Doblas, A., Sánchez-Ortiga, E., Saavedra, G., and Huang, Y.-P. (2015b). Fast axial-scanning widefield microscopy with constant magnification and resolution. *Journal of Display Technology*, 11(11):913–920.

[Mendiburu, 2012] Mendiburu, B. (2012). *3D movie making: stereoscopic digital cinema from script to screen*. CRC Press.

[Mendlovic, 2013] Mendlovic, D. (2013). Toward a super imaging system [invited]. *Applied Optics*, 52(4):561–566.

[Mousnier et al., 2015] Mousnier, A., Vural, E., and Guillemot, C. (2015). Partial light field tomographic reconstruction from a fixed-camera focal stack. *arXiv preprint arXiv:1503.01903*.

[Nagahara et al., 2010] Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H., and Nayar, S. K. (2010). Programmable aperture camera using lcos. In *European Conference on Computer Vision*, pages 337–350. Springer.

[Nayar and Nakagawa, 1994] Nayar, S. K. and Nakagawa, Y. (1994). Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831.

[Ng, 2006] Ng, R. (2006). *Digital light field photography*. PhD thesis, Stanford University.

[Ng et al., 2005] Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. (2005). Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11.

[Nugent, 1987] Nugent, K. A. (1987). Coded aperture imaging: a Fourier space analysis. *Applied Optics*, 26(3):563–569.

[Orth and Crozier, 2012] Orth, A. and Crozier, K. (2012). Microscopy with microlens arrays: high throughput, high resolution and light-field imaging. *Optics Express*, 20(12):13522–13531.

[Pajares and De La Cruz, 2004] Pajares, G. and De La Cruz, J. M. (2004). A wavelet-based image fusion tutorial. *Pattern Recognition*, 37(9):1855–1872.

[Park et al., 2008] Park, J.-H., Baasantseren, G., Kim, N., Park, G., Kang, J.-M., and Lee, B. (2008). View image generation in perspective and orthographic projection geometry based on integral imaging. *Optics Express*, 16(12):8800–8813.

[Pastoor and Wöpking, 1997] Pastoor, S. and Wöpking, M. (1997). 3-D displays: A review of current technologies. *Displays*, 17(2):100–110.

[Patwary and Preza, 2015] Patwary, N. and Preza, C. (2015). Image restoration for three-dimensional fluorescence microscopy using an orthonormal basis for efficient representation of depth-variant point-spread functions. *Biomedical Optics Express*, 6(10):3826–3841.

[Pertuz et al., 2013] Pertuz, S., Puig, D., Garcia, M. A., and Fusiello, A. (2013). Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images. *IEEE Transactions on Image Processing*, 22(3):1242–1251.

[Preza and Conchello, 2004] Preza, C. and Conchello, J.-A. (2004). Depth-variant maximum-likelihood restoration for three-dimensional fluorescence microscopy. *JOSA A*, 21(9):1593–1601.

[Preza et al., 1992] Preza, C., Miller, M. I., Thomas, L. J., and McNally, J. G. (1992). Regularized linear method for reconstruction of three-dimensional microscopic objects from optical sections. *JOSA A*, 9(2):219–228.

[Ray, 2015] Ray, S. (2015). *Scientific photography and applied imaging*. CRC Press.

[Russ, 2016] Russ, J. C. (2016). *The Image Processing Handbook*. CRC press.

[Sarder and Nehorai, 2006] Sarder, P. and Nehorai, A. (2006). Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Processing Magazine*, 23(3):32–45.

[Schechner and Kiryati, 2000] Schechner, Y. Y. and Kiryati, N. (2000). Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162.

[Schechner et al., 2000] Schechner, Y. Y., Kiryati, N., and Basri, R. (2000). Separation of transparent layers using focus. *International Journal of Computer Vision*, 39(1):25–39.

[Schild et al., 2012] Schild, J., LaViola, J., and Masuch, M. (2012). Understanding user experience in stereoscopic 3d games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 89–98. ACM.

[Schuhladen et al., 2013] Schuhladen, S., Petsch, S., Liebetraut, P., Müller, P., and Zappe, H. (2013). Miniaturized tunable imaging system inspired by the human eye. *Optics Letters*, 38(20):3991–3994.

[Schwiegerling and Tyo, 2013] Schwiegerling, J. and Tyo, J. S. (2013). Relating transverse ray error and light fields in plenoptic camera images. In *SPIE Optical Engineering+ Applications*, pages 884203–884203. International Society for Optics and Photonics.

[Sellent and Favaro, 2014] Sellent, A. and Favaro, P. (2014). Which side of the focal plane are you on? In *Computational Photography (ICCP), 2014 IEEE International Conference on*, pages 1–8. IEEE.

[Smith and Collar, 2012] Smith, M. D. and Collar, B. T. (2012). Perception of size and shape in stereoscopic 3d imagery. In *IS&T/SPIE Electronic Imaging*, pages 82881O–82881O. International Society for Optics and Photonics.

[Subbarao, 1990] Subbarao, M. (1990). On the depth information in the point spread function of a defocused optical system. *Computer Vision Laboratory, Department of Electrical Engineering State University of New York, Stony Brook, NY*, pages 11794–2350.

[Subbarao and Surya, 1994] Subbarao, M. and Surya, G. (1994). Depth from defocus: a spatial domain approach. *International Journal of Computer Vision*, 13(3):271–294.

[Surman, 2013] Surman, P. (2013). Stereoscopic and autostereoscopic displays. In *3D-TV System with Depth-Image-Based Rendering*, pages 375–411. Springer.

[Tian et al., 2011] Tian, J., Chen, L., Ma, L., and Yu, W. (2011). Multi-focus image fusion using a bilateral gradient-based sharpness criterion. *Optics Communications*, 284(1):80–87.

[Tsai and Yeh, 2010] Tsai, C. G. and Yeh, J. A. (2010). Circular dielectric liquid iris. *Optics Letters*, 35(14):2484–2486.

[van Beurden et al., 2012] van Beurden, M. H., IJsselsteijn, W., and Juola, J. (2012). Effectiveness of stereoscopic displays in medicine: a review. *3D Research*, 3(1):1–13.

[Veeraraghavan et al., 2007] Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., and Tumblin, J. (2007). Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69.

[Watanabe and Nayar, 1995] Watanabe, M. and Nayar, S. K. (1995). Telecentric optics for constant magnification imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19:1360–1365.

[Wu et al., 2012] Wu, Q., Guo, S., Ma, Y., Gao, F., Yang, C., Yang, M., Yu, X., Zhang, X., Rupp, R. A., and Xu, J. (2012). Optical refocusing three-dimensional wide-field fluorescence lifetime imaging microscopy. *Optics Express*, 20(2):960–965.

[Xiao et al., 2013] Xiao, X., Javidi, B., Martinez-Corral, M., and Stern, A. (2013). Advances in three-dimensional integral imaging: sensing, display, and applications [invited]. *Applied Optics*, 52(4):546–560.

[Xu et al., 2015] Xu, M., Ren, H., and Lin, Y.-H. (2015). Electrically actuated liquid iris. *Optics Letters*, 40(5):831–834.

[Yang and Li, 2010] Yang, B. and Li, S. (2010). Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59(4):884–892.

[Zhang and Guo, 2009] Zhang, Q. and Guo, B.-l. (2009). Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Processing*, 89(7):1334–1346.

[Zheng et al., 2011] Zheng, G., Kolner, C., and Yang, C. (2011). Microscopy refocusing and dark-field imaging by using a simple LED array. *Optics Letters*, 36(20):3987–3989.

[Zhou et al., 2011] Zhou, C., Lin, S., and Nayar, S. K. (2011). Coded aperture pairs for depth from defocus and defocus deblurring. *International Journal of Computer Vision*, 93(1):53–72.

[Zhou et al., 2012] Zhou, C., Miau, D., and Nayar, S. K. (2012). Focal sweep camera for space-time refocusing. *Dept. Comput. Sci., Columbia Univ., Tech. Rep. CUCS-021-12*.

[Zhou and Nayar, 2011] Zhou, C. and Nayar, S. K. (2011). Computational cameras: Convergence of optics and processing. *IEEE Transactions on Image Processing*, 20(12):3322–3340.

[Zhou et al., 2014] Zhou, Z., Li, S., and Wang, B. (2014). Multi-scale weighted gradient-based fusion for multi-focus images. *Information Fusion*, 20:60–72.