
DISSERTATION

Master in Electrical and Electronic Engineering

# Improved Image Rendering for Focused Plenoptic Cameras with Extended Depth-of-field

José Nunes dos Santos Filipe

Leiria, January 2019


DISSERTATION

Master in Electrical and Electronic Engineering

# Improved Image Rendering for Focused Plenoptic Cameras with Extended Depth-of-field

José Nunes dos Santos Filipe

Master dissertation performed under the supervision of Professor Luís Miguel Oliveira Pegado de Noronha e Távora and Professor Sérgio Manuel Maciel de Faria, both from Escola Superior de Tecnologia e Gestão of Instituto Politécnico de Leiria.

Leiria, January 2019

*I pass with relief from the tossing*
*sea of Cause and Theory to the*
*firm ground of Result and Fact.*

***(Winston Churchill)***

# Acknowledgements

I would like to start by thanking to my advisers Prof. Luís Miguel de Oliveira Pegado de Noronha e Távora and Prof. Sérgio Manuel Maciel de Faria, that guided me and helped me during this journey. I also express my gratitude towards all members of the DermoPLeno research group, namely Prof. Pedro António Amado Assunção, Prof. Rui Fonseca-Pinto, Rui Miguel Leonel Lourenço, João Miguel Pereira da Silva Santos, Pedro Miguel Marques Pereira for the insight provided through our many debates and the knowledge sharing. Additionally, I would like to thank João Filipe Monteiro Carreira and Ricardo Jorge Santos Monteiro for the help provided when any doubt, either theoretical or practical, rose up.

I would like to thank the opportunity for working as a researcher in the project "Dermo-Pleno – Dermo-Plenoptic imaging for skin lesion assessment" (DermoPleno/2016) and for the financial support provided by Instituto de Telecomunicações, through Fundação para a Ciência e Tecnologia. I also would like to acknowledge the lab facilities and research environment provided by Instituto de Telecomunicações and Escola de Tecnologia e Gestão of Instituto Politécnico de Leiria.

Finally, I would like to thank my parents, José Pereira dos Santos Filipe and Elsa Maria de Jesus Nunes Filipe, for their support. I would also like to give a very special thanks to my grandmother, Maria do Rosário Fernandes Pereira. Additionally, I would also like to express my gratitude towards my friends, that either directly or indirectly helped me during this journey.

# Abstract

This dissertation presents a research work on rendering images from light fields captured with a focused plenoptic camera with extended depth of field.

A basic overview of the 7 dimensional plenoptic function is first given, followed by a description of the Two-Plane Parametrisation. Some of the various methods used for sampling the plenoptic functions are then described, namely those equivalent to acquisition functions implemented by the camera gantry, the unfocused plenoptic camera and the focused plenoptic camera. State-of-the-art image rendering algorithms have also been studied both for focused and unfocused plenoptic cameras.

A comprehensive study of the behaviour of focus metrics when applied to images rendered form a focused plenoptic camera is presented, including 34 of the most widely used metrics in the literature. Due to high frequency artefacts, caused by the rendering process, it was found that the currently available focus metrics yield inflated values for this kind of images, leading to misindication, where worse-focused images have better focus measures. Subjective tests were carried out, in order to corroborate these results.

Then, methods for minimizing the rendering artefacts are proposed. An algorithm for choosing the maximum patch size for each micro-image was designed, in order to minimize the distortions caused by the vignetting effect of the micro-lens. Then an inpainting algorithm, based on anisotropic diffusion inpainting, is used to minimize the remaining artefacts present in the borders between adjacent micro-images.

Finally, a method to deal with the redundant information generated by a plenoptic camera with extended depth of field is presented. Three different views of the same scene are rendered, with the three different types of lenses. Then, it is proven that making any linear combination of the images always results in worse focus than selecting the better focused one. Thus, a multi-focus image fusion algorithm is proposed to merge the three images captured by a extended depth-of-field camera into a single one, which presents higher focus level than any of the three individual images.

***Keywords:*** Light Field, Plenoptic, Rendering, Al-in-focus, Focus, Focus Metrics, Subjective Quality Assessment, Image Inpainting, PDEs, Image Fusion, Image Registration

# Resumo

Esta dissertação apresenta um trabalho de investigação sobre síntese de imagens capturadas com uma câmara, plenóptica focada com profundidade de campo estendida.

Uma descrição geral dos fundamentos da função plenóptica de 7 dimensões é apresentada, seguida de uma descrição da Parametrização de Dois Planos. Alguns dos vários métodos usados para amostragem da função plenóptica são descritos em seguida, nomeadamente métodos equivalentes à função de aquisição implementada por um pórtico de câmaras, uma câmara plenóptica desfocada e uma câmara plenóptica focada. Alguns algoritmos do estado da arte em síntese deste tipo de imagens são também descritos, quer para a câmara plenóptica focada, quer para a câmara plenóptica desfocada.

Apresenta-se também um estudo do comportamento de métricas de foco, aplicadas a imagens sintetizadas a partir de uma câmara plenóptica focada. Um conjunto de 34 das métricas mais usadas na literatura foi usado neste estudo. Devido aos artefactos presentes nas imagens, causados pelo processo de síntese, foi concluído que estas métricas apresentam valores inflacionados para este tipo de imagens, indicando erradamente imagens mais mal focadas, como tendo valores de focagem melhores. Também foram conduzidos testes subjectivos, de forma a corroborar os resultados obtidos.

Seguidamente, um método para lidar com a informação redundante gerada pelas câmeras plenópticas com profundidade de campo estendida é apresentado. Três vistas diferentes da mesma cena são sintetizadas, com os três tipos de lentes diferentes. Em seguida, foi provado que fazer uma qualquer combinação linear das imagens é sempre pior, no que toca ao foco, do que escolher a imagem mais focada. Assim, um algoritmo de fusão de imagens multi-foco é usado para fundir as três imagens numa, que apresenta um nível de foco maior do que qualquer uma das três imagens iniciais.

***Palavras-chave:*** Light Field, Plenoptico, Síntese, Tudo-em-foco, Foco, Métricas de Foco, Avaliação de Qualidade Subjectiva,*Inpainting* de Imagens, EDPs, Fusão de Imagens, Registo de Imagens

# Contents

**Acknowledgements** **iii**

**Abstract** **v**

**Resumo** **vii**

**Contents** **x**

**List of Figures** **xi**

**List of Tables** **xiii**

**1 Introduction** **1**

1.1 Context and Motivation . . . . . 1

1.2 Objectives . . . . . 2

1.3 Outline . . . . . 2

**2 Light Fields – State of the Art** **3**

2.1 The Plenoptic Function . . . . . 4

2.2 Parametrisation of the Plenoptic Function . . . . . 4

2.3 Unfocused Plenoptic Camera . . . . . 7

2.3.1 Rendering Method . . . . . 8

2.4 Focused Plenoptic Camera . . . . . 10

2.4.1 Rendering Method . . . . . 12

**3 Measuring Focus in Images Rendered from Focused Plenoptic Cameras** **19**

3.1 Optimum versus non-optimum extended depth-of-field reconstruction . . . 20

3.2 Evaluation methodology . . . . . 22

3.2.1 Objective metrics . . . . . 22

3.2.2 Subjective evaluation . . . . . 24

3.3 Results . . . . . 25

3.3.1 Objective Measures . . . 25
3.3.2 Subjective Results . . . 27
3.3.3 Statistical Analysis and Relationship Between the Objective and Subjective Results . . . 27
3.4 Final Remarks . . . 33

**4 Method to Improve the Rendering Process** **35**
4.1 Shortcomings of the rendering process . . . 35
4.1.1 Low light-intensity image regions . . . 35
4.1.2 Patch border artefacts . . . 36
4.2 Proposed Method to Improve the Rendering . . . 37
4.2.1 Finding the maximum patch size . . . 37
4.2.2 Rendering — patch selection . . . 38
4.2.3 Patch Border Anisotropic Inpainting . . . 39
4.3 Results and Discussion . . . 43
4.4 Final Remarks . . . 46

**5 Combining All-in-Focus Images from Extended Depth of Field Cameras** **47**
5.1 Integrating the Lens Types into the Rendering . . . 47
5.2 Merging the Information from the 3 Types of Lens . . . 51
5.2.1 Image Registration . . . 51
5.2.2 Multi-Focus Image Fusion . . . 53
5.3 Results . . . 54
5.4 Final Remarks . . . 55

**6 Conclusions** **57**

**Bibliography** **59**

**A Contributions** **67**

# List of Figures


2.1 Graphical representation of the Two-plane Parametrisation (adapted from [20]). . . . . 5
2.2 Matrix of views resultant of the Two-plane Parametrisation (from [23]). . . 6
2.3 Formation of an Epipolar Plane Image (EPI) (from [23]). . . . . 6
2.4 A - Simple Camera setup; B - Plenoptic Camera setup; C - Physical correspondence with the Two-plane Parametrisation (adapted from [1]). . . . . 8
2.5 Illustration of the unfocused plenoptic camera view rendering method. . . . 9
2.6 adapted from [15]. . . . . 10
2.7 Schematic representation of a plenoptic camera 1.0 (a), a Keplerian focused plenoptic camera (b) and a Galilean focused plenoptic camera (c) (from [29]). 11
2.8 Lenslet acquired with an unfocused camera (at the left) and a focused (at the right) plenoptic camera, respectively (from [30]). . . . . 12
2.9 Graphical representation of the simple refocus method, described in [7]. . . 13
2.10 Representation of the rendering process with constant patch size, discussed in [7,31]. Big patches originate repeated structures, small patches originate missing structures and only patches with a correct size are able to render an image without discontinuities (adapted from [31]). . . . . 14
2.11 Result of a refocused image from a focused plenoptic camera with hexagonal patches. Two overlapping sheets of paper at different depths are shown. The sheet on the background is focused. The sheet on the foreground shows repeated structures. . . . . 15
2.12 Representation of the extended depth-of-field rendering method, as described in [8]. . . . . 17

3.1 Representation of artificial edges between patches (even when the optimum patch size is selected). . . . . 21
3.2 Some of the images used throughout this study are shown. Images with optimal patch size ($I$) on the left and image rendered with patches one pixel bigger then the optimum patch size ($I_{+1}$), on the right. . . . . 23

3.3 Zoom-in of an extended depth-of-field reconstructed image (a) using optimum patch size; (b) using a patch size 1 pixel larger than the optimum . . 28
3.4 Subjective evaluation - preference of observers (%). . . . . 28
3.5 Preference of observers (%), with the confidence intervals at 95%. Test image A presents a confidence interval of ±22.4, image B of ±22.4, image C of ±27.4, image D of ±29.6, image E of ±25.1 and image F of ±27.4. . . 29

4.1 Artefacts inherent to the rendering algorithm [8] (Image A). . . . . 36
4.2 Proposed method including pre-processing (maximum patch size) and anisotropic inpainting of patch borders. . . . . 37
4.3 Schematic showing the process for computing the maximum patch size for a micro-image (Equation 4.2). . . . . 38
4.4 Image 4.4a shows a view rendered with normalization of the patch size, and image 4.4b shows a view rendered without the normalization step. . . . . 39
4.5 Generalisation of the inpainting problem: depiction of the know region $\phi$, the region to be inpainted $\Omega$ and boundary $\delta\Omega$ between the two (from [72]). 40
4.6 Application of isotropic diffusion (Figure 4.6b) and anisotropic diffusion (Figure 4.6c) for inpainting the gap between the white bars on Image 4.6a. 42
4.7 Images from Fig. 4.1 rendered with the proposed method (Image A). . . . 44
4.8 Rendered image (Image B). . . . . 45
4.9 Rendered image (Image F). . . . . 45

5.1 Representation of the lenslet structure typical of the focused plenoptic cameras with extended depth of field, where the 3 types of lenses with different focal lengths can be clearly seen[1]. . . . . 48
5.2 Unfocused camera-like acquisition system, simplified in 2-dimensions. . . . 49
5.3 Plot of the Gauss error function. . . . . 51

# List of Tables

3.1 Relative variation of the focus metrics, $\Delta m(\%)$, assessed in images rendered with a patch size 1 pixel larger than optimum. Variations of the cross border "quality indicator", $\Delta q(\%)$, are also presented. . . . . 26

3.2 Evaluation of the Jaccard-Needham [61] and Rogers-Tanimoto [62] Indices, to compare the evaluations of the focus metrics with the subjective preferences. . . . . 32

4.1 Smoothness gain, $G_q$ of the proposed method for six images (A to F). . . . 44

5.1 Assessment of the focus improvement. . . . . 55

# Chapter 1

# Introduction

This chapter presents a brief introduction to the work developed in the scope of this dissertation. Firstly, the context and motivation are laid out, secondly the objectives of the research are discussed and finally the structure of the dissertation is presented.

## 1.1 Context and Motivation

In recent years, the increasing commercial availability of plenoptic cameras, namely of handheld light field cameras, as well as the increment in computational power, resulted in an growing interest for the extended capabilities offered by this type of photography. Such capabilities include refocusing images, change the point of view of the camera, render all-in-focus images and acquire depth maps of a scene. All of these capabilities are possible after a plenoptic photograph has been acquired, unlike traditional photography, where the point of view and the focus plane can only be changed before an image is taken.

These features makes light field photography suitable for a wide range of applications, from some more entertaining ones, such as sports photography [1], where the refocus feature is important, since the photographer does not need to worry with focus when taking the picture, and film industry, such as the view point changing in the Time Bullet effect in the Matrix movies [1], to industrial and research applications, such as light field microscopy [2], depth and image restoration [3], particle velocimetry [4], quality inspection [5], and medical imaging [6], among others.

## 1.2 Objectives

A crucial step in all applications described above is the view rendering process. Furthermore, the spatial resolution of rendered views from a plenoptic camera of the first generation (unfocused plenoptic camera) are low compared with what current commercial cameras can achieve. As such, this dissertation concentrates on studying and improving view rendering methods for focused plenoptic cameras, mainly by studying the methods described in [7,8], where patches of the lenslet image are tiled together in order to render refocused and all-in-focus images, analyse the problems exhibited by these methods, and finally propose some solutions.

## 1.3 Outline

In order to do so, Chapter 2 presents some theoretical background as well as the description of the state of the art. Chapter 3 explores the measurement of focus level in views rendered with this type of cameras. Chapter 4 describes the improvements made to the rendering pipeline. Chapter 5 explores a methodology to efficiently merge the redundant information given by focused plenoptic cameras with extended depth-of-field, and finally Chapter 6 presents the conclusions of this work.

# Chapter 2

# Light Fields – State of the Art

The concept of light field representation of the three-dimensional visual space defined around an arbitrary observer has been introduced more than a century ago [9]. However, only recently the creation and commercialization of hand-held cameras brought the subject back to the spotlight.

A crucial part of modern plenoptic photography is the rendering of multiple views, allowing refocus, perspective shifts and even tridimensional reconstruction. These characteristics make these cameras suitable for quite diverse applications, such as medical imaging, industrial inspection, museology, video-vigilance, etc.

One of the earliest approaches to Light Field imaging was done by the French Nobel Prize winner Gabriel Lippmann [10], around 1908. Under the name of Integral Photography, Lippmann developed a device that was similar to a 4 by 3 grid of analog cameras, allegedly based on insects eyes, which allowed the capture and visualization of tridimensional perspectives of the photographed scene. These concepts were further studied by Ives [11], and in 1936, by the Russian physicist Andrey Gershun, who coined the term ''Light Field" [12], among others. In 1968, Chutjian [13] developed the first digital light field camera and in 1991, Adelson first referred to light field cameras as Plenoptic cameras [14]. Already in the 21st century, more precisely in 2006, Ng built a hand-held plenoptic camera [15] and Levoy [16] adapted the light field technology to microscopy. In 2008, Lumsdaine and Georgiev [7] improved the spatial resolution of plenoptic cameras, introducing the focused plenoptic camera, or plenoptic camera 2.0. A more detailed historical revision of the evolution of light field photography can be founds in [17]. This chapter presents a brief review of the state-of-the-art in light field technology. First, the theoretical concepts are laid out and explored, followed by a discussion about the physical construction of such cameras. Then, methods for rendering views from light field data are discussed.

## 2.1 The Plenoptic Function

The plenoptic function is a theoretical mathematical model that describes the intensity distribution of light at every point in space and time [18]. As such, this is a function of 7 variables, as shown in Equation 2.1, where $\phi$ and $\theta$ indicate the orientation of the light rays in spherical coordinates, $\lambda$ refers to the wavelength, $t$ refers to the temporal dimension and the triplet $V_x$, $V_y$ and $V_z$ identify the Cartesian position of the observer.

$$P(\theta, \phi, \lambda, t, V_x, V_y, V_z) \tag{2.1}$$

According to [18], a description of every photograph would be implicitly contained in such representation of light. Nonetheless, due to the complexity and the high dimensionality of the plenoptic function, it is not possible to get a complete plenoptic representation of a scene, since it is impossible o look simultaneously from every point of view, for every wavelength and moment in time. In spite of this, it is still possible to sample the plenoptic function, by probing the desired scene from different points of view, filter the wavelength to the traditional RGB colour space and sampling time at regular frame intervals. This ultimately allows the use of processing techniques that extends the capabilities of common photography, such as obtaining the depth map of the scene, refocus and all in focus images.

## 2.2 Parametrisation of the Plenoptic Function

As stated in Section 2.1, the plenoptic function is a 7D model, which renders it cumbersome and difficult to work with. Nonetheless, as explained in [19], it is possible to reduce the dimensionality of the plenoptic functions, by making few assumptions. First of all, as the object of study is plenoptic photography, and photographs are static in time, it is possible to drop the time dependence or reduce this time dependence to discrete sampling instead, in the case of video signals. In order to further simplify the function, only the luminance shall be considered, which drops the wavelength as a continuous parameter. In reality, the wavelength information is encoded in the RGB information provided by the cameras, therefore discretising this dimension does not cause significant practical loss.

At this stage, the simplified plenoptic function only depends on the position of the observer and on the orientation of the light rays, $P(\theta, \phi, V_x, V_y, V_z)$, which is a 5D function. In order to further reduce the plenoptic function dimensionality, light should by analysed in terms of rays, under the assumption of the absence of occlusions and that all the light emitters

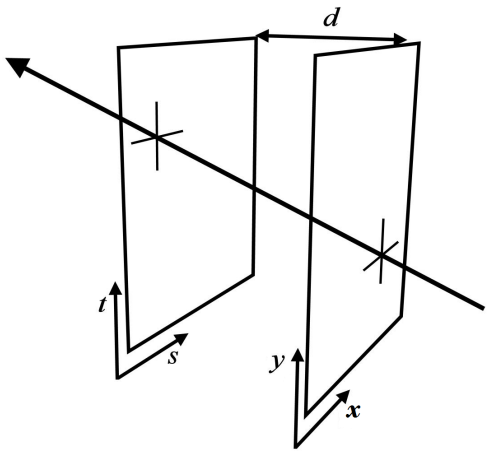



Figure 2.1: Graphical representation of the Two-plane Parametrisation (adapted from [20]).

that constitute the scene are non-Lambertian surfaces[1]. If the luminance of a light ray emitted by one point from a scene is measured, then it will be constant, no matter how far from the scene the measurement is taken [22]. To further illustrate this concept, it is possible to imagine a laser beam in free space pointing towards infinity. The luminance emitted by the laser source would be the same, no matter how far from the laser source it is measured. Therefore, if the $V_xV_y$ plane is parallel to the scene, the plenoptic function representing that same scene is independent of $V_z$. This implies that the position of the observer can be represented with only two meaningful spatial coordinates, $V_x$ and $V_y$, since the distance from the scene does not change how light rays arrive to the observer.

The plenoptic function is now reduced to 4 dimensions. Nonetheless, it is common to use the Two-Plane Parametrisation to represent light fields, instead of the spatial and directional coordinates used so far. This parametrisation consists on defining two arbitrary parallel planes, separated by a fixed known distance, $d$. This way, a light ray is defined in this parametrisation by the two sets of 2D coordinates that locate the points where the light rays intersect each of the planes, as shown in Figure 2.1. Now, it is possible to arrange these two planes such that one plane corresponds to a matrix of views, captured from different perspectives, and the other plane corresponds to the coordinates within each view, as shown in Figure 2.2. Usually the coordinates $(s, t)$ are assigned to the plane of the matrix of views and the $(x, y)$ are the coordinates inside each view. Equation 2.1 is thus reduced to:

$$P(s, t, x, y) \tag{2.2}$$

[1]Lambertian surfaces are surfaces that present Lambertian reflectance, *i. e.*, the surface reflects the received light equally in all possible viewing directions, meaning that the light is reflected in an isotropic manner [21]. *Ergo*, the brightness of a Lambertian object appears to be the same, regardless of the observer's position.

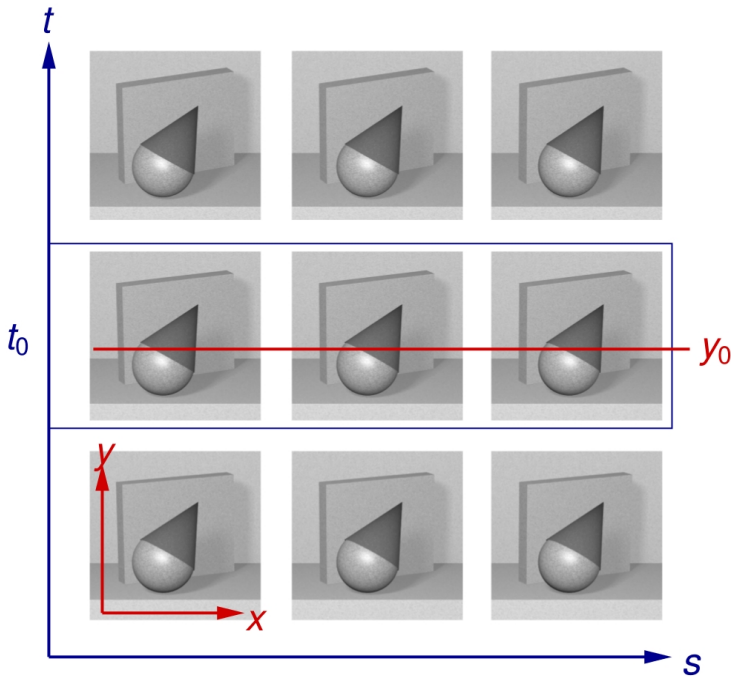



Figure 2.2: Matrix of views resultant of the Two-plane Parametrisation (from [23]).

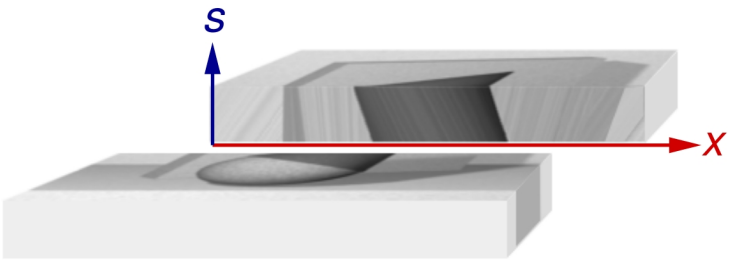



Figure 2.3: Formation of an Epipolar Plane Image (EPI) (from [23]).

Using this parametrisation, that was first introduced in [19], if a pair $(s,t)$ is chosen, $P(s,t,x,y)$ yields a view of the scene from a selected $(s_0,t_0)$ point of view. If pair $(x,y)$ is chosen, $P(s,t,x,y)$ yields the pixel $(x_0,y_0)$ from every single view from the matrix of views. Consequently, if a set of $(s,t,x,y)$ coordinates is chosen, this function yields the $(x_0,y_0)$ pixel from the view $(s_0,t_0)$.

Moreover, these type of images can also be parametrised as a set of Epipolar Plane Images, or EPIs [24], by selecting a pair $(s_0,x_0)$ or $(t_0,y_0)$. To further analyse its meaning, lets consider that a pair $(t_0,y_0)$ has been chosen. Fixing $t_0$ is equivalent to select a row of views from the matrix of views. Fixing $y_o$ is equivalent to select a single row of pixels within each of the views. Putting it together, by fixing a par $(t_0,y_0)$, we are choosing a single row of pixels $y_0$ from each view that belongs to the row $t_0$ from the matrix of views, as shown in Figure 2.2. This can also be seen as stacking the views from the row $t_0$ following the order of the $s$ axis and cutting the pile along $y_0$, as shown in Figure 2.3. This originates an image with vertical dimension, equal to the number of the views

that the matrix of views has along $t_0$, and a horizontal dimension, equal to the horizontal resolution of each view.

To further understand the EPI geometry, lets consider a regular photographic camera, moving linearly along $t_0$ and at a constant speed, and taking photos at constant time intervals (that is, considering $s$ as the time axis). Then, the EPI shows how the structures represented within the $y_0$ line appear to move from the point of view of the camera – it is fundamentally a representation of disparity. Considering an object that is represented within the $y_0$ line in every view belonging to $t_0$, that is very far from the moving camera (one may say infinitely far) it does not appear to change from the point of view of the camera. In other words, the position of the object represented in the $y_0$ line remains constant across all views that belong to $t_0$, which creates a perfectly vertical structure in the EPI. As the object approaches the camera, its position appears to change more and more between contiguous views — an effect known as parallax — which generates structures with increasing slopes in the EPI. This implies that, if it is possible to measure the slopes of the structures of every EPI, then it is possible to reconstruct the three-dimensional representation of the scene [15, 25], *i. e.* the EPI contains 3D information of the scene.

## 2.3 Unfocused Plenoptic Camera

One way to sample the plenoptic function of a scene relies on Equation 2.2 and the Two-plane Parametrisation. It consists on using a matrix of cameras to capture different and evenly spaced views, as done in [26], for instance. A variation of this approach consists on using a moving camera to capture views from various perspectives, which is equivalent to use a matrix of cameras. This setup is usually called light field gantry.

Nonetheless, such devices require much space and are not portable at all. To overcome these problems, Ng [15] proposed the addition of an array of micro-lenses between the main lens and the photo-sensor of a simple camera (Figure 2.4 - A), such that the scene is focused on the plane of the micro-lenses, as shown in Figure 2.4 - B. Thus, the photo-sensor is able to capture slightly different positions of what would be a single pixel in a simple camera. This way, the projection of a micro-lens onto the photo-sensor is called micro-image, and the set of all micro-images projected by a micro-lens array (MLA) is called a lenslet image.

Physically, the $xy$ plane is located at the main lens, since the position at which a light ray crosses the main lens, determines the position of the information carried by that ray within a generated view. The physical location of the $st$ plane however, is not so

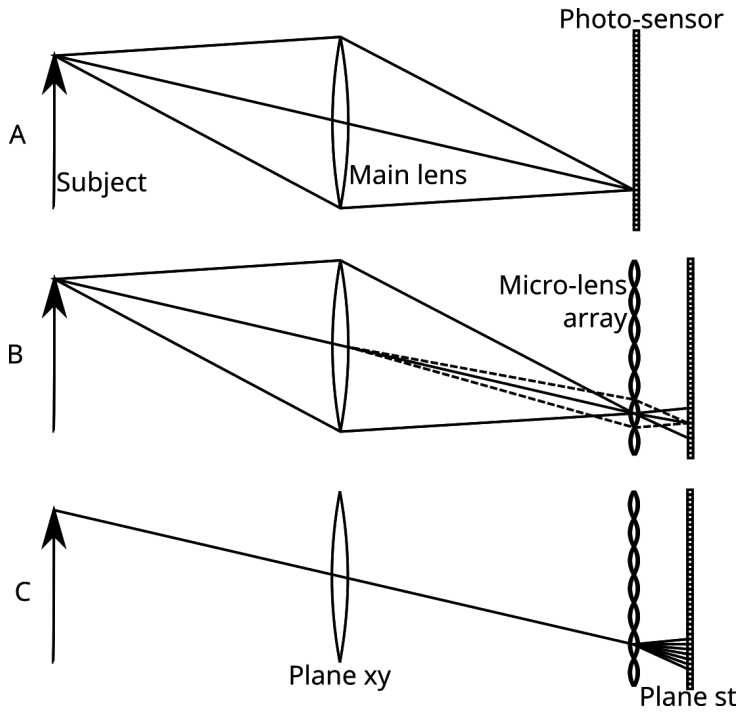



Figure 2.4: A - Simple Camera setup; B - Plenoptic Camera setup; C - Physical correspondence with the Two-plane Parametrisation (adapted from [1]).

clear. Given that the information of that light ray that crossed the $xy$ plane is going to be refracted by the MLA, representations of the same information in slightly different positions inside neighbour micro-images are originated. Since that, in order to render different views, different sections of the micro-images are used, the $st$ is located at the MLA.

### 2.3.1 Rendering Method

In light of the description of this kind of camera the first micro-lens of the upper left corner of the micro-lens array originates all the first pixels from the upper left corner of each view. The second micro-lens originates all the second pixels, and so on. So, rendering the central view from the MLA, essentially consists in choosing the central pixel of the $s$ by $t$ pixels structure originated by the first micro-lens, and then concatenate it with the central pixel of the second micro-lens, and so on. To render the view one pixel to the right of the central view, simply choose the pixel on the right side of the central pixel from all micro-lenses and concatenate them. This process is illustrated in Figure 2.5, where the thick boxes on the right represent the boundaries of the micro-images, the thin line squares represent the pixels, and the colours represent the selected pixels form each micro-lenses. It can be noticed that the order of the colours do not change form the lenslet representation on the right, to the rendered view to the left, *i.e.*, the concatenations of the

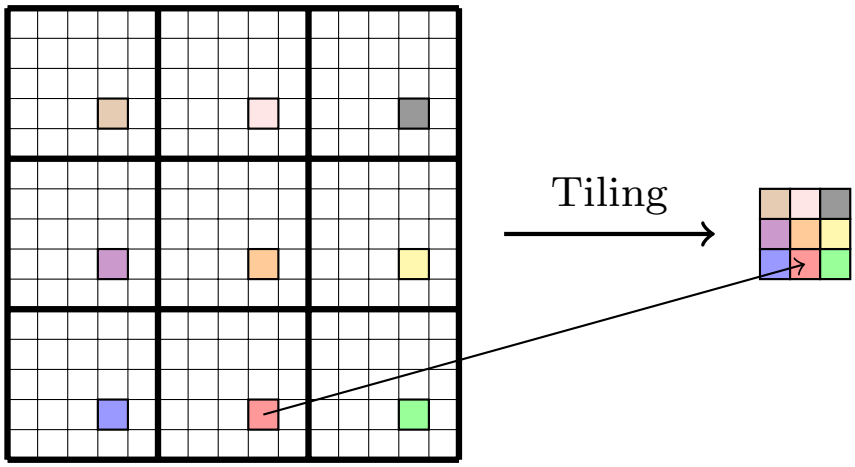



Figure 2.5: Illustration of the unfocused plenoptic camera view rendering method.

pixels is done keeping the spatial order of the pixels. In this image it is also possible to observe that the spatial resolution of the rendered view is much smaller than the spatial resolution of the lenslet image. Then, it is possible to easily obtain a matrix of views, and therefore a 4D representation of a light field.

However, another advantage of light field photography is the possibility to render views focused in different planes (known as refocus), or even an all focus view (also known as image with extended depth of field), after the light field has been captured. In order to achieve this effect, the Two-plane Parametrization is modelled according with the equation of the irradiance of an image [27], as seen in Equation 2.3, where $F$ denotes the distance between the two planes ($st$ and $xy$), $A(s,t)$ represents the aperture function for each micro-lens (yields the value 1 for full aperture) [28] and $\theta$ is the angle of incidence of the ray $(s,t,x,y)$ with the plane $st$.

$$E(x,y) = \frac{1}{F^2} \iint P(s,t,x,y) \cdot A(s,t) \cdot \cos^4\theta \, ds \, dt \tag{2.3}$$

Furthermore, according to [28], the constant $1/F^2$ can be ignored and, under the paraxial approximation (*i.e.*, considering that angle $\theta$ is small, which, approximated by the Taylor series, results in $\cos\theta \approx 1$), $\cos^4\theta \approx 1$. Additionally, for the propose of refocusing, the full aperture of the micro-lens is used, resulting in $A(s,t) = 1$. Taking these considerations into account, it is possible to rewrite Equation 2.3 in a simplified manner, as shown in Equation 2.4.

$$\bar{E}(x,y) = \iint P(s,t,x,y) \, ds \, dt \tag{2.4}$$

Given this parametrization, refocusing is simple a matter of "relocating" the $x,y$ plane (or

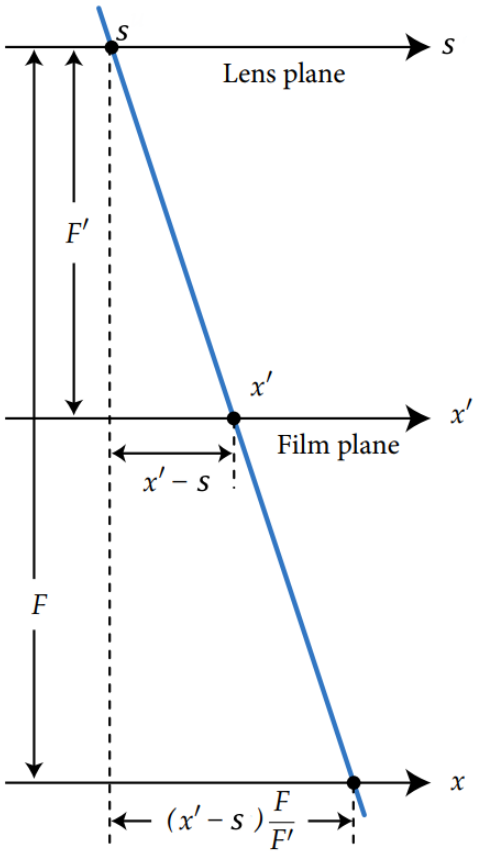


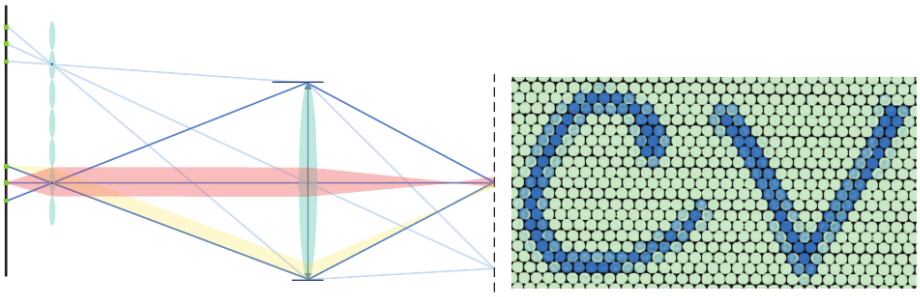

Figure 2.6: adapted from [15].

film plane), in order to coincide with the desired focus plane. In other words, refocusing consists in changing the distance $F$. Figure 2.6 shows a 2D simplification of this problem. Originally, the film plane $x$ is located at a distance $F$ from the lens plane $s$. However, to refocus the image to film plane $x'$, at a distance $F'$ from the lens plane, $x$ needs to be expressed in terms of $x'$. Nonetheless, form Figure 2.6, it is easy to conclude that, by similar triangles, $x = s + \frac{x'-s}{\alpha}$, where $\alpha = F/F'$. Analogously, the same argument can be used with the other pair of axis, yielding $y = t + \frac{y'-t}{\alpha}$. By replacing these expressions in Equation 2.4, yielding the refocusing equation ( Equation 2.5), where the ratio $\alpha$ defines the location of the focus plane.

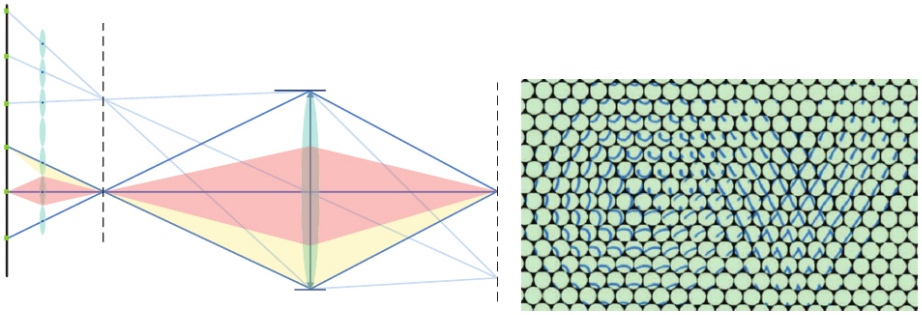
$$\bar{E}(x', y'; \alpha) = \iint P(s, t, s + \frac{x' - s}{\alpha}, t + \frac{y' - t}{\alpha})\, ds\, dt \tag{2.5}$$

## 2.4 Focused Plenoptic Camera

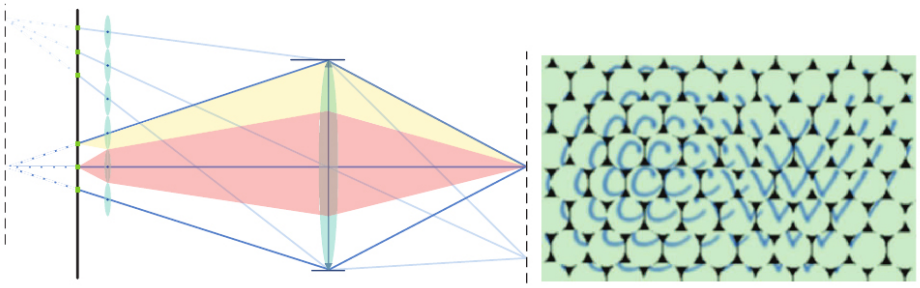
The type of camera described in Section 2.3, presents a constrain: the spatial resolution of each view is limited to the number of micro-lens in the MLA, since one micro-image



(a) Plenoptic camera



(b) Focused plenoptic camera in telescopic case



(c) Focused plenoptic camera in binocular case

Figure 2.7: Schematic representation of a plenoptic camera 1.0 (a), a Keplerian focused plenoptic camera (b) and a Galilean focused plenoptic camera (c) (from [29]).

provides only one pixel to each view. This results in views with poor spatial resolution which is undesirable for some applications.

To overcome this problem the authors in [7] propose to change the focus plane such that the MLA behaves as an array of pinhole cameras with narrow field of view. This change allows the use of more pixels from each micro-image to render each view, which ultimately results in views with greater spatial resolution when compared to the previous camera type. These devices are known as Focused Plenoptic Camera, or Plenoptic Camera 2.0. However, there are two different camera designs proposed by [7] able achieve this. One is known as telescopic or Keplerian focused plenoptic cameras and the design change consists in rearranging the the distance between the main lens and the MLA in such a way that focus plane falls before the MLA, such that the light rays reaching the MLA can be re-focused into the sensor, as shown by the red beam in Figure 2.7b. As can also be seen in

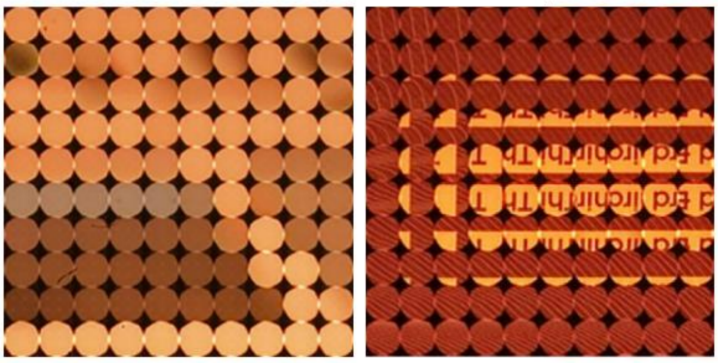


Figure 2.8: Lenslet acquired with an unfocused camera (at the left) and a focused (at the right) plenoptic camera, respectively (from [30]).

the figure, this type of cameras produce inverted micro-images, due to the inversion of the capture image that occurs after the light rays cross the location of the virtual focus plane. However, the MLA (and consequently the photo-sensor), can be placed before the plane. This way, the MLA can re-focus the light rays into the photo-sensor, without the image being formed, as can be seen in Figure 2.7c. These cameras are known as Galilean or binocular plenoptic cameras. This approach produces micro-images that are not inverted.

Nonetheless, the relationship between the Two-plane Parametrisation and the physical configuration of the camera is no longer straightforward and the rendering algorithms (which shall be addressed in the following sections) are considerably more complex.

From Figure 2.8, it is possible to observe that the micro-lenses of the unfocused plenoptic camera capture a *macropixel*, *i.e.*, a set of pixels from each view, which looks blurred, whilst the micro-lens of the focused plenoptic camera contain sharp portions (focused) of the full resolution image. This means that, for focused plenoptic cameras, the micro-images must be focused themselves, in order to obtain a focused image. This is due to the fact that a portion of the micro-image (*i. e.* a patch) is used for rendering views, hence that patch must be focused in order to produce an image in focus. Thus, focused plenoptic cameras lose the depth of field characteristic of the unfocused plenoptic cameras, in order to gain a much higher spatial resolution. To increase the depth of field offered by focused plenoptic cameras it is suggested in [13] to use several types of micro-lens, each one focused at a different focus plane. This means that one type of micro-lens can be used when the remaining are out of focus.

### 2.4.1 Rendering Method

The methods described in Section 2.3.1 do not apply to the focused plenoptic camera, since there is not a straightforward relation between the Two-plane Parametrisation and the

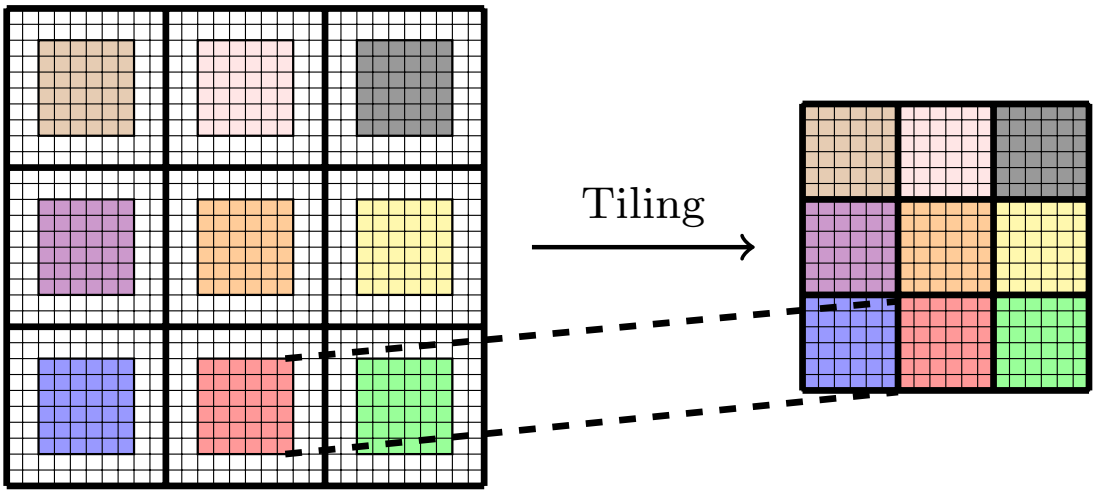
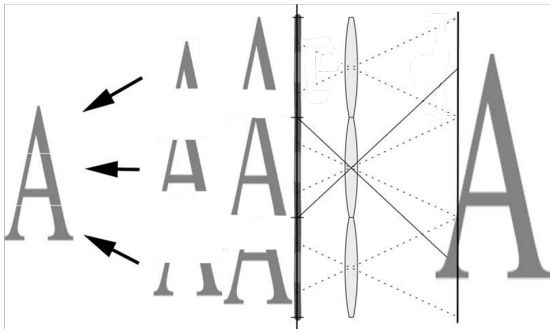



Figure 2.9: Graphical representation of the simple refocus method, described in [7].

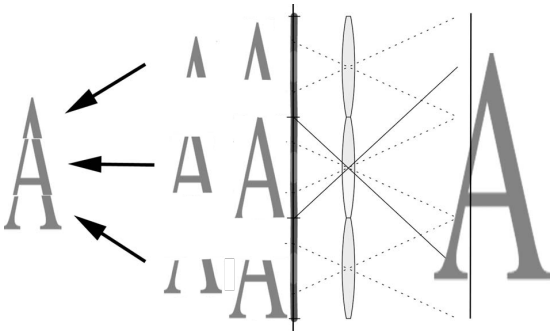
model of the camera, as mentioned in Section 2.4. One pipeline to render refocused images is described in [31]. First of all, to render an image from a focused plenoptic camera, a patch of pixels from each micro-image is selected and all patches are concatenated side-by-side. In this case, if the centre of the selected patch is the same as the centre of the micro-image, then the rendered view is the centre view. If the centre of the patch is shifted one pixel to the right of the centre of the micro-image, then the rendered view is shifted one pixel to the right, in relation to the centre view, as illustrated in Figure 2.9. Similarly to Figure 2.5, the thick lines at the right represent the boundaries of the micro-lens and the thin ones, delimit the pixels. In this image it is possible to notice that a patch of pixels (represented by the coloured regions within the micro-lenses) is selected from each micro-lenses and tiled together, keeping the same order. It is also possible to see that the spatial resolution of the rendered view is much higher than the spatial resolution of the views rendered form unfocused plenoptic cameras (see Figure 2.5).

To shift the focus plane, the size of the patch is set, making it the same for all patches in all micro-images. This way, the patches that contain the objects in focus will be perfectly matched together (Figure 2.10a), whilst patches that contain objects beyond the focus plane will not match, showing incomplete objects with missing pixels (Figure 2.10c). Patches containing objects on closer planes will not match either, but in this case, some repeated structures appear in the reconstructed image, showing repeated pixels (Figure 2.10b), as shown in Figure 2.11. Figure 2.10 depicts how this effects manifest in the rendered image.

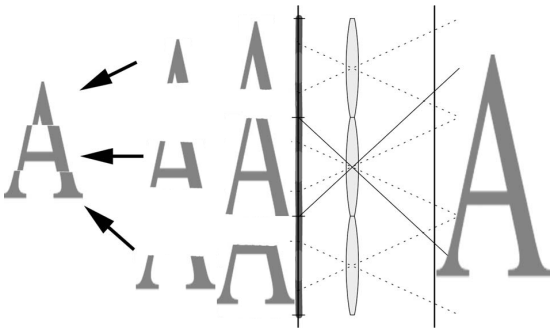
From this behaviour, it is possible to conclude that choosing patches with a smaller size allows to render images that are focused at closer planes relative to the camera, while using patches with larger size results in focus in planes further away from the camera. However, this method produces unnatural looking views, because the repeated or missing



(a) In focus.



(b) Patch size too large.



(c) Patch size too small.

Figure 2.10: Representation of the rendering process with constant patch size, discussed in [7,31]. Big patches originate repeated structures, small patches originate missing structures and only patches with a correct size are able to render an image without discontinuities (adapted from [31]).

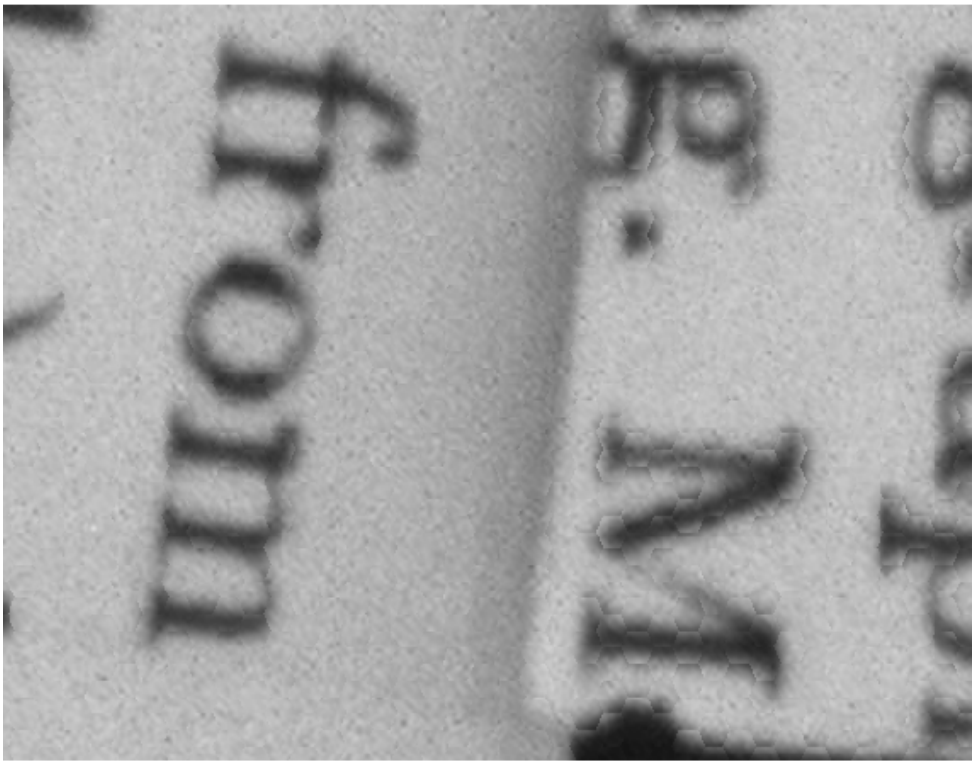


Figure 2.11: Result of a refocused image from a focused plenoptic camera with hexagonal patches. Two overlapping sheets of paper at different depths are shown. The sheet on the background is focused. The sheet on the foreground shows repeated structures.

structures create artificial and pronounced discontinuities in the reconstructed images, seriously harming the subjective quality of the rendered images. To solve this problem, it is suggested in [31] that a 'blending' effect can be achieved by averaging the same spatial point across all the micro-images where such point is present.

Since a generic visual scene contains elements at different depths, rendering an extended (up to full) depth-of-field image involves the use of patches with different sizes. Therefore, in order to obtain extended depth-of-field images it is necessary to resize all different patches to a given dimension before tiling them together. In earlier approaches [7,31], the optimum patch size was obtained by using information from overlapping areas in adjacent micro-images. To simplify the rendering process for all-in-focus images, Wanner *et al.* [8] make use of the assumption that natural images are, in general, piecewise continuous and such assumption still holds, even when transitioning across different neighbouring patches, given that they are in focus. The process starts by rendering a focus stack *i. e.*, a set of images focused at different planes, by the method previously described. Let this set be represented by:

$$\Gamma = \{\Omega_m \mid m = m_{min}, ..., m_{max}\} \tag{2.6}$$

In Equation 2.6, $m$ denotes the patch size, which in turn defines the position of the focus plane. The value of $m_{min}$ corresponds to the size of the smallest patch, while $m_{max}$ is equal to the size of the micro-images. Given that a patch should contain a portion of the image to be rendered, the choice of $m_{min}$ is important, since it can not be so small that the content of the patch is lost. A Laplacian filter ($\nabla^2$) is applied to each image, in order to emphasize regions of rapid intensity change. Then, the gradients for the patch $p_{i,j}$ in each filtered micro-image $w(i,j)$ are evaluated, by integrating along its borders, $b(p_{i,j})$, considering only directions that are perpendicular to its edges. For all images $\Omega_m$, the sum of those orthogonal gradients is obtained by integrating along its borderline, as given by Equation 2.7

$$\sigma(m,i,j) = \oint_{b(p_{i,j})} \vec{\nabla}\left(\nabla^2 \Omega_m\right) \cdot \hat{n_b}, \qquad \Omega_m \in \Gamma \tag{2.7}$$

In Equation 2.7, $\hat{n_b}$ denotes a unit vector perpendicular to border $b$. The optimum patch size is obtained by finding the patch with the smoothest possible transitions across borders. For such purpose, a map $z(i,j)$ containing the optimum patch sizes, *i.e.*, those that minimize the sharpness of transitions between patch $p_{i,j}$ and its neighbours, is found through minimisation, according to Equation 2.8.

$$z(i,j) = \underset{m}{\operatorname{argmin}}\ \sigma(m,i,j), \quad \forall\, i,j \tag{2.8}$$

The extended depth-of-field image, $\Omega_{ez}$, is then rendered by tiling together the $p_{i,j}$ patches with optimum sizes $z(i,j)$. As mentioned before, the patches must be resized to the same spatial dimension to allow concatenation, as can be seen in Figure 2.12. Similarly, to Figure 2.5 and Figure 2.9, the thick lines on the right represent the boundaries of the micro-lens and the thin ones delimit the pixels. In the image, it is possible to notice that a patch of pixels, represented by the coloured regions within the micro-lenses, is selected from each micro-lenses. Since they originally present quite distinct patch sizes, they need to be upsampled to a common resolution and tiled together, keeping the same spatial order.

Finally, in order to concatenate all patches with optimal (and obviously different) sizes together, it is necessary to upsample them all to the same size, which is, usually, the size of the largest patch possible. At last, the patches are concatenated to form a view with extended depth of field.

This algorithm was developed for focused plenoptic cameras, but it can be used for cameras with extended depth of field as well, if each type of micro-lens is considered as a different plenoptic camera. In this case, it is possible to obtain a number of central all focus

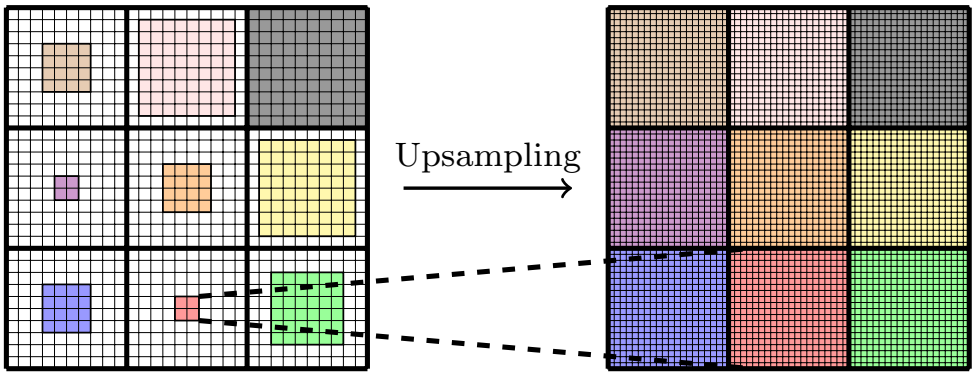



Figure 2.12: Representation of the extended depth-of-field rendering method, as described in [8].

views equal to the number of micro-lens' types. These different views show a slight offset between them, due to the physical placement of the micro-lens. As suggested in [8], it is possible to take advantage of these characteristics and use the cross-correlation to find and correct the offset between the images and then take its average. In spite of being a simple method to combine the information of all lens, it does not take full advantage of the fact that the images of each type of lens are focused at different planes.

# Chapter 3

# Measuring Focus in Images Rendered from Focused Plenoptic Cameras

In general, a good focus level of an image, or image region, is achieved when fine details are perceived and distinguished from each other. In computational terms, this can be measured by using focus metrics that estimate the image sharpness and provide some numerical measure of the focus level. This means that such metrics are designed to capture the presence of sharp transitions and high contrast, represented by high gradients in the spatial domain and high frequencies in the transform domain. Since the magnitude of focus levels is highly dependent on the image content, metrics are only comparable when computed for different (yet similar) images of the same visual scene (*i.e.*, the same content), though with different acquisition parameters. A common characteristic of all focus metrics is the monotonic relationship between the value provided by the focus measure (*i.e.*, the focus level) and the visual focus perceived in the images. Accordingly, the higher the focus level, the more focused is the image, or region, under observation [32].

The computational methods used in plenoptic image refocusing optimise the focus level based on the optical characteristics of the acquisition system. Therefore, it would be expected that focus metrics provided maximum values for the focus level of extended depth-of-field reconstructed images (*i.e.*, all in-focus), which are optimally focused across the whole image area. However, such direct relationship does not hold for the commonly used patch-based extended depth-of-field reconstruction methods described in Chapter 2. To the author's best knowledge, this has not been addressed in previous studies. Thus, a comprehensive study on the use of a large number of different focus metrics (*i.e.*, 34 in total) in the evaluation of extended depth-of-field images reconstructed from light fields is presented in this chapter. It is shown that the best refocused images, from the computational point of view are not those which provide higher focus levels. This is also

confirmed through subjective evaluation, where the best perceived images were found to be those reconstructed with the optimum patches, but not those with higher objective focus level.

As described in Chapter 2, contiguous micro-images may include objects located at different depths on the scene, thus different micro-images may contribute with patches of different sizes. Therefore, in order to obtain extended depth-of-field images it is necessary to upsample patches of different sizes to a common fixed size, prior to tile them together. The best patch size is determined from the overlapping area between neighbour micro-images, which presents a degree of dependence on the disparity in that region of the scene.

## 3.1 Optimum versus non-optimum extended depth-of-field reconstruction

If extended depth-of-field reconstruction is used to obtain all-in-focus images, then such images are optimally focused across the whole visual scene and their sharpness (*i.e.* focus level), measured by common focus metrics, is expected to be maximum. In opposition, when non-optimum patch sizes are used in the image reconstruction (*i.e.,* resulting images are not in focus), the magnitude of its focus level is expected to be lower than that obtained for images reconstructed with optimum patch sizes. However, given the intrinsic characteristics of the patch-based method used in extended depth-of-field reconstruction, such behaviour does not always occur, since non-optimum patch sizes originate discontinuities in the images, wrongly inflating the magnitude of such focus metrics.

In fact, even though the optimum patch sizes found by the rendering algorithm produce the smoothest transitions across patch borders, it is likely that some artificial discontinuities still appear along the matching borders due to local texture variations, as it can be seen in Figure 3.1. If such discontinuities are relevant enough to affect the focus level computed by commonly used metrics, then no valid conclusions can be drawn from such focus measurements. This is due to the fact that a higher focus level may be mostly due to such artificial transitions, not from the image visual content itself.

To investigate this problem, it is necessary to define an objective measure for the smoothness of patch transitions. For this purpose, a parameter $q$ has been defined based on Equation 2.7, as can be seen in Equation 3.1, where higher values of $q$ means sharper

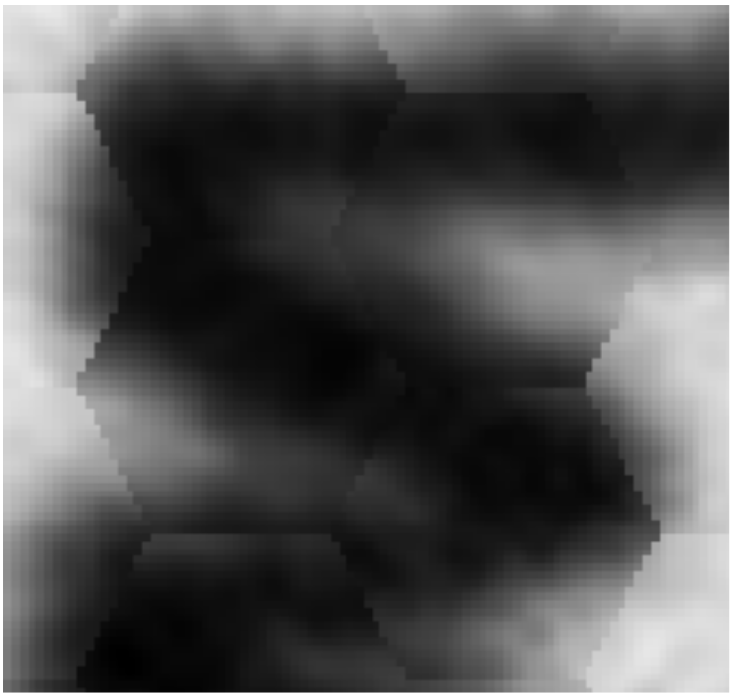


Figure 3.1: Representation of artificial edges between patches (even when the optimum patch size is selected).

border transitions.

$$q = \sum_{i,j} \oint_{b(p_{i,j})} \vec{\nabla}(\nabla^2 I(x,y)) \cdot \hat{n_b} \tag{3.1}$$

Like in Equation 2.7, Equation 3.1 measures the sharpness of the transitions between adjacent micro-images, by computing the gradient of the Laplacian on the whole refocused image, only in the perpendicular directions of each patch border.

In the light of the previous discussion, and since the extended depth-of-field reconstruction method has low impact at the patch borders, $q$ might be used as a relative no-reference quality indicator for the specific type of impairments under study, *i.e.*, smoothness across patch borders. According to the definition given by Equation 3.1, a higher value of $q$ means lower quality due to the presence of sharper transitions exist across patch borders.

The following sections describe the study carried out in order to evaluate the impact of this effect in the classical focus metrics available in the literature. In Section 3.2, the application of 34 metrics on 6 all-in-focus images, rendered with optimum patch size, is compared with the application of the same metrics on the same light fields, but with the 6 images rendered with patches one pixel larger than the optimum patch size, both in height and width, aiming to evaluate if the studied metrics present inflated values, when applied to these latter images. Furthermore, a subjective study carried out in order to

access if the results given by the objective metrics are according to the perceived opinion of the observers, is also described.

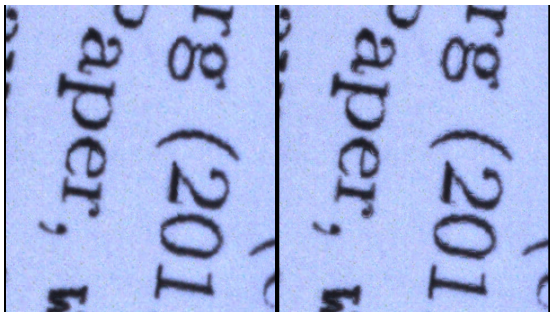
## 3.2 Evaluation methodology

Given the problem identified in Section 3.1, the following step is to assess if the inflation of the focus metrics is occurring in the metrics available in the literature. The methodology defined to carry out such study is described as follows: a set of 6 light fields were captured with a focused plenoptic camera. The depth of field of the captured light fields was kept narrow, in order to preserve the focus within each micro-image. Some of the 6 images used are displayed in Figure 3.2. Two medical images are included in the set, as they represent an example where focus accuracy is important. These images were acquired with a focused plenoptic camera *Raytrix* R42, f/2.8 Micro Lens Array, UHR optics c-mount (25mm, f/1.8), placed at distance of 140mm from object (between the main lens and the scene, approximately 20mm between the camera and the scene) and with natural white colour LED illumination).

A comparative study was carried out using optimal patch sizes (*i.e.*, all-in-focus images) versus non-optimal patch sizes. The former images (denoted as $I$) were reconstructed using the base algorithm described in Section 2.4.1, while the latter (denoted as $I_{+1}$) were reconstructed using patch sizes one pixel larger than the optimum. The performance of 34 objective focus metrics was evaluated by comparing the corresponding focus level obtained for the two types of reconstructed images, $I$ and $I_{+1}$. Then an equivalent comparative evaluation was carried out through subjective assessment. The reconstructed images were also characterised in terms of the quality indicator $q$, defined by Equation 3.1.

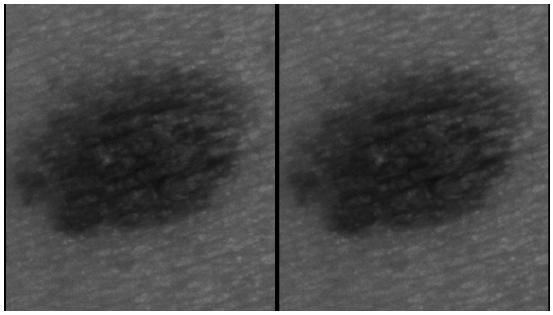
### 3.2.1 Objective metrics

Focus metrics are defined as functions, which should exhibit the following characteristics. They must be monotonic, such that two similar yet slightly different images do not produce the same output; unimodal, meaning that the function yields only one extreme (usually a maximum) corresponding to the best-focused image; insensitive to noise and with good variability, which allows a clear differentiation between focused and unfocused images. In this study, 34 different focus measures available from the current literature, with the above characteristics, were considered. The majority of focus metrics can be classified either as spatial-domain metrics, when applied directly to the image pixels, or transform-based metrics, when some kind of mathematical transform is used. Each of these two categories can be further grouped into several subcategories, according to spe-

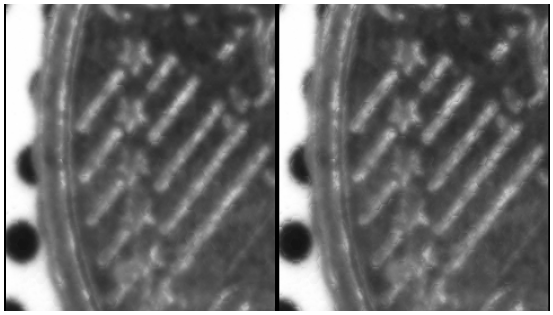


(a) Image A. Written piece of paper.

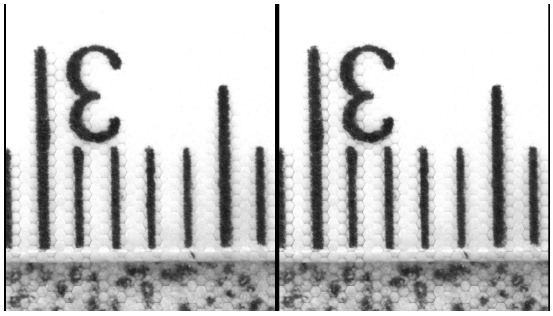
(b) Image B. Printed circuit board (Arduino).



(c) Image C. Skin spots (nevus).



(d) Image E. 50 euro cents coin.



(e) Image F. Ruler scale (cm).

Figure 3.2: Some of the images used throughout this study are shown. Images with optimal patch size ($I$) on the left and image rendered with patches one pixel bigger then the optimum patch size ($I_{+1}$), on the right.

cific mathematical operators used to compute the focus measures. This work uses several metrics of each category, covering a wide range of different methods, which also allows relevant characterization of state-of-the-art algorithms. The spatial-domain metrics used in this work are based on the following methods:

- **Gradient**: based on the norm of the gradient vector, composed by the directional partial first derivatives of the image. The following methods were considered: Gradient energy [32], Gaussian derivative [33], Thresholded absolute gradient [34], Squared gradient [35], Tenengrad [36] and Tenengrad variance [37];
- **Laplacian**: based on the use of the norm of the Laplacian, composed by the directional partial second derivatives of the image. This study considered the following methods: Energy of Laplacian [32], Modified Laplacian [38], Diagonal Laplacian [39] and Variance of Laplacian [37];
- **Statistics**: based on statistical indicators extracted from the images, such as histograms, mean, median and variance. The used methods were: Absolute central moments [40], Chebyshev moments [41], Eigenvalues [42], Gray-level variance [36], Gray-level local variance [37], Normalized grey-level variance [34], Modified grey-level variance [43], Histogram entropy [36] and Histogram range [44];
- **Miscellaneous**: algorithms based on several different mathematical tools. The following were considered the most representative: Brenner's measure [34], Image contrast [45], Image curvature [46], Helmli and Scherer's mean [46], Steerable filter [47], Spatial frequency measure [35], Vollath's standard deviation [34] and Vollath's autocorrelation [34].

The following transform-based metrics were used in this work:

- **DCT**: based on the discrete cosine transform. The methods selected for this study are the following: DCT energy ratio [48], DCT reduced energy ratio [49] and Modified DCT [49];
- **Wavelet**: based on the wavelet transform, such as the following: Sum of wavelet coefficients [50], Variance of wavelet coefficients [50], Ratio of the wavelet coefficients [51] and Ratio of curvelet coefficients [52].

### 3.2.2 Subjective evaluation

The subjective evaluation methodology was based on the proposal of the document ITU-R BT.500-13 [53]. For comparison of subjective scores and objective focus measure, the same

two reconstructed images, evaluated with the objective focus metrics, were presented side-by-side to the observers. One of these images was rendered with the algorithm described in Section 2.4.1 ($I$) and the other one was rendered with patch sizes one pixel larger than the optimal ones ($I_{+1}$). These two images were displayed side-by-side on a LG 55UJ63 55' monitor, each one occupying half of the horizontal resolution of the monitor. Since the resolution of the images (4460 x 3470) was much higher than the full-HD display resolution (1920 x 1080), only a limited region of each image (with full HD resolution: 1920x1080) was shown to the observers, thus avoiding zooming out. The subjective evaluation was conducted individually after a training session. One pair of images was presented to each observer during 10 seconds and the observer could review it as many times as needed. The presentation order of each image pair and position on the display were randomized for each observer. Overall, 18 non-expert observers evaluated the 6 images of the data set. All of them passed the Snellen vision test for visual acuity [54]. The observers were asked to choose the image with higher perceived quality, *i.e.*, less artefacts.

## 3.3 Results

### 3.3.1 Objective Measures

The objective results obtained from the 34 focus metrics are shown in Table 3.1 for the 6 pairs of reconstructed images (A to F). These figures were calculated as the relative focus level obtained between the images reconstructed with optimum patch sizes plus one pixel ($I_{+1}$) and those reconstructed with the optimum patch sizes ($I$). Denoting by $m(\cdot)$ an absolute focus level obtained from any particular metric, the relative focus level $\Delta m$ shown in Table 3.1 is defined according to Equation 3.2.

$$\Delta m = \frac{m(I_{+1}) - m(I)}{m(I)} \times 100 \tag{3.2}$$

When $\Delta m$ is positive, it means that the non-optimum patch sizes produce an image with higher focus level than that obtained with optimum patch sizes.

The relative variation of the "quality indicator" $q$, defined in Equation 3.1, is also shown in Table 3.1 for each pair of reconstructed images. However, as like the focus level $m$, $q$ is also presented as a variation, $\Delta q$, which is defined by Equation 3.3.

$$\Delta q = \frac{q(I_{+1}) - q(I)}{q(I)} \times 100 \tag{3.3}$$

| | | Focus Metrics \ Images | A $\Delta q = 5.37$ | B $\Delta q = 11.61$ | C $\Delta q = 1.04$ | D $\Delta q = 2.20$ | E $\Delta q = 5.44$ | F $\Delta q = 10.56$ |
|---|---|---|---|---|---|---|---|---|
| Spatial Domain | Gradient | Gradient Energy | 11.51 | 9.77 | 6.23 | 8.24 | 3.37 | 12.05 |
| | | Gaussian Derivative | 1.85 | 2.86 | -1.89 | -0.37 | 4.35 | 5.83 |
| | | Thresholded Abs. Grad. | 7.18 | 5.44 | 4.96 | 6.97 | 7.84 | 6.09 |
| | | Squared Gradient | 9.39 | 9.06 | 6.24 | 9.68 | 3.45 | 9.35 |
| | | Tenengrad | 9.65 | 7.51 | 6.19 | 10.70 | 6.72 | 11.72 |
| | | Tenengrad Variance | 7.94 | 13.13 | 6.68 | 14.36 | 0.20 | 28.40 |
| | Laplacian | Energy of Laplacian | 31.41 | 15.78 | 15.71 | 19.56 | 8.69 | 27.29 |
| | | Modified Laplacian | 12.06 | 7.58 | 8.12 | 9.50 | 9.55 | 9.80 |
| | | Diagonal Laplacian | 12.37 | 7.92 | 8.25 | 10.27 | 10.19 | 9.72 |
| | | Variance of Laplacian | 31.41 | 15.78 | 15.71 | 19.56 | 8.69 | 27.29 |
| | Statistics | Chebyshev Moments | 0.22 | 0.98 | 0.01 | -0.05 | 0.25 | -2.16 |
| | | Absolute Central Mom. | -0.11 | -1.34 | -0.73 | -0.02 | 0.05 | -0.31 |
| | | Eigenvalues | 0.02 | -0.13 | 0.16 | 0.00 | -0.45 | 2.22 |
| | | Gray-level Variance | -0.12 | -2.08 | -1.22 | 0.11 | -0.15 | 1.60 |
| | | Gray-level Local Variance | 3.75 | 2.93 | -2.18 | 1.36 | 4.62 | 1.53 |
| | | Norm. Gray-level Local Var. | 0.00 | -0.75 | -0.49 | 0.14 | -0.20 | 1.92 |
| | | Modified Gray-level Var. | 10.76 | 7.87 | 2.88 | 6.28 | 9.49 | 12.32 |
| | | Histogram Entropy | 0.04 | -0.17 | -0.14 | 0.03 | 0.00 | -0.01 |
| | | Histogram Range | 0.00 | 0.00 | 0.46 | -0.46 | 0.00 | 0.00 |
| | Miscellaneous | Brenner's Measure | 11.16 | 10.25 | 7.48 | 12.19 | 7.85 | 16.48 |
| | | Image Contrast | 7.15 | 5.57 | 4.93 | 7.03 | 7.96 | 6.08 |
| | | Image Curvature | 0.01 | -0.97 | -0.64 | 0.08 | 0.12 | -0.23 |
| | | Helmli and Scherer's Mean | 0.29 | 1.15 | 0.08 | 0.09 | 0.19 | 0.22 |
| | | Steerable Filter | 0.13 | 2.82 | -1.12 | -0.64 | 1.99 | 4.25 |
| | | Spatial Frequency Measure | 7.19 | 5.51 | 4.98 | 6.98 | 7.92 | 6.13 |
| | | Vollath's Autocorrelation | 11.24 | 8.21 | 6.68 | 11.36 | 10.58 | 11.62 |
| | | Vollath's Standard Dev. | -0.22 | -2.73 | -1.45 | -0.05 | 0.05 | -0.48 |
| Transforms Domain | DCT | DCT Energy Ratio | 4.75 | 3.12 | 2.19 | 2.43 | 3.37 | 4.70 |
| | | DCT Reduced Energy Ratio | 3.46 | 2.33 | 0.91 | 1.18 | 3.07 | 3.51 |
| | | Modified DCT | 39.02 | -8.84 | -51.98 | 13.62 | 0.00 | -39.27 |
| | Wavelets | Sum of Wavelet Coeffs. | 13.16 | 8.68 | 8.47 | 9.60 | 10.65 | 11.62 |
| | | Var. of Wavelet Coeffs. | 15.86 | 7.59 | 7.88 | 10.94 | 4.07 | 12.82 |
| | | Ratio of Wavelet Coeffs. | 34.72 | 17.93 | 19.27 | 20.53 | 8.38 | 28.97 |
| | | Ratio of Curvelet Coeffs. | -0.65 | 9.71 | 11.92 | 7.06 | 3.48 | -14.04 |

Table 3.1: Relative variation of the focus metrics, $\Delta m(\%)$, assessed in images rendered with a patch size 1 pixel larger than optimum. Variations of the cross border "quality indicator", $\Delta q(\%)$, are also presented.

For $\Delta q$ positive, the non-optimum patch sizes produce lower quality (*i.e.*, stronger patch border transitions) than that obtained with the optimum patch sizes.

A relevant observation in the results of Table 3.1 is that almost all metrics exhibit a misleading behaviour: *i.e.,* an indication of better focus ($\Delta m > 0$), while in fact images $I_{+1}$ exhibit higher artificial border transition gradients ($\Delta q > 0$) and non-optimal focus due to the larger patch sizes. Therefore, for the majority of focus metrics and image pairs, evaluation of image focus through comparison of focus levels leads to the wrong conclusion that using non-optimal patch sizes yields better focused images. This effect is to be associated to the influence of the border transitions, not the visual content itself. This can be confirmed by zooming in relevant regions of reconstructed images, as shown in Figures 3.3a and 3.3b. Even though the relative focus level indicate that Figure 3.3b is shaper than Figure 3.3a, this is due to the stronger patch border transitions introduced by the use of non-optimal patch sizes, not because of better focus.

Only in a very limited number of cases the results differ from the behaviour discussed

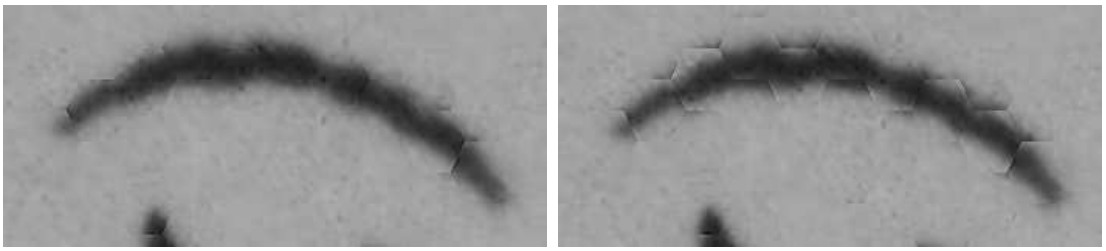
above. In the case of statistical based metrics, these results are likely to be associated to the fact that the image mean value might be affected by the presence of artificial edges. It is also worth to observe that transform-domain metrics, in particular Wavelet based methods, are those with a stronger indication of an erroneous better focus. Furthermore and most importantly, is to notice that widely used methods, like the Energy of Laplacian or the Energy of Gradient, are amongst those that consistently fail to produce higher focus level for the better focused images, because $\Delta m$ is always positive.

### 3.3.2 Subjective Results

The subjective results, presented in Figure 3.4, confirm that reconstructed images with optimum patch sizes are preferred in the majority of tests. This figure shows percentage of test cases where the observers have chosen either $I$ or $I_{+1}$ as their preferred image in each of the 6 tested image pairs. For all tested images, at least 61% of the inquired observers have chosen $I$ over $I_{+1}$. In fact, in images with a wider range of depth, as images A and B, this percentage rises to 78%, which is consistent with the high $\Delta q$ values obtained for these two images, when compared to the others. On average, in 70.5% (STD = 6.8%) of the tests the observers preferred the images $I$ (reconstructed with optimum patch sizes) over $I_{+1}$. For confidence intervals that keep the expected value above 50%, the results shown in Figure 3.4 present a certainty level above 87%, with the exception of image D, where this value drops to 68%. This is due to the fact that this image depicts human skin, where the artificial edges are much harder to notice, because of the inherent texture irregularities.

### 3.3.3 Statistical Analysis and Relationship Between the Objective and Subjective Results

Figure 3.5, shows the same results represented in Figure 3.4, but with the standard 95% confidence intervals. These confidence intervals acknowledge the possibility of the true preference of the observers be less than 50%, namely for images C, D, E and F. At the first sight, this seems to indicate that these results are inconclusive. However, by dealing with the outliers and by modelling the voting process, it is possible to narrow the confidence intervals and extract more accurate results form the subjective data. This was proposed in [55] (and further explored in [56]). The core idea is to model the raw preference of the observer, according to Equation 3.4. Considering the $e$-th tested image, such that $e = 1, ..., E$, evaluated by the $s$-th observer, such that $S = 1, ..., S$, in this model, the raw preferences are modelled as a random variable $X_{e,s}$, composed by $x_e$, which represents the true preference and $B_{e,s}$, which is the random part of $X_{e,s}$, representing the variable



(a) (b)

Figure 3.3: Zoom-in of an extended depth-of-field reconstructed image (a) using optimum patch size; (b) using a patch size 1 pixel larger than the optimum

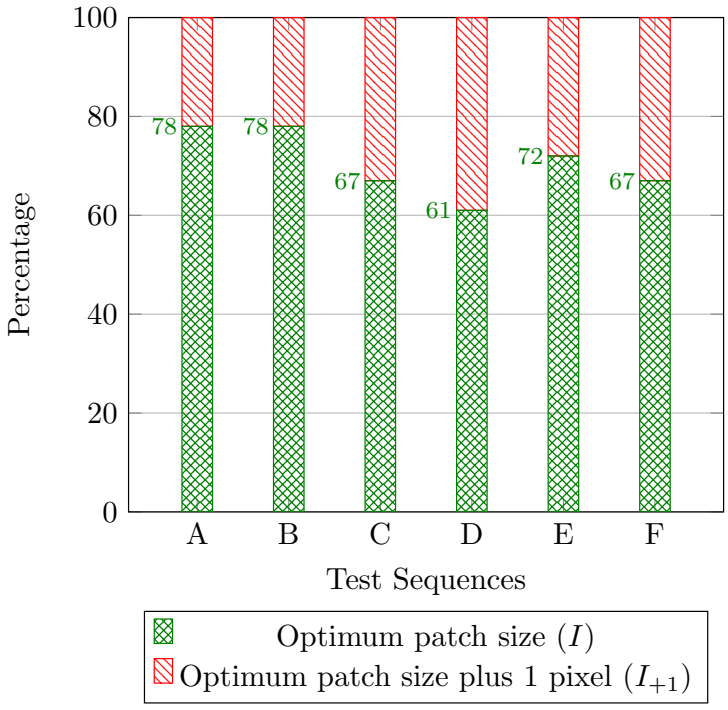



Figure 3.4: Subjective evaluation - preference of observers (%).

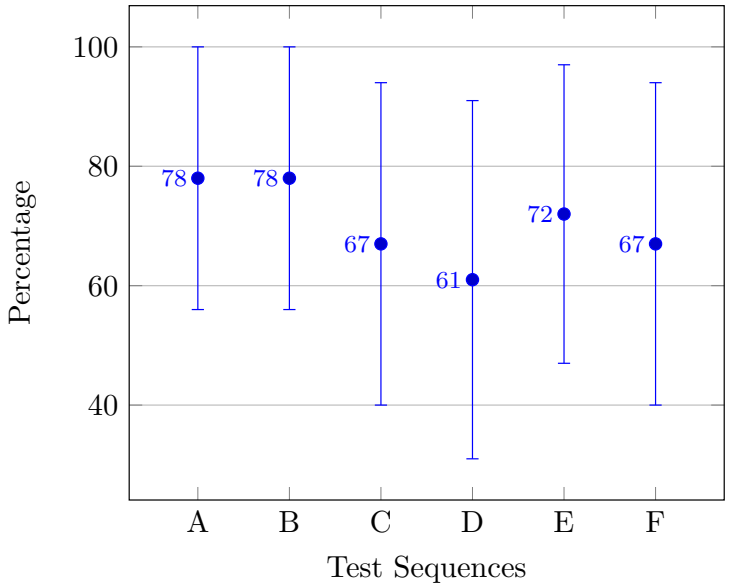



Figure 3.5: Preference of observers (%), with the confidence intervals at 95%. Test image A presents a confidence interval of ±22.4, image B of ±22.4, image C of ±27.4, image D of ±29.6, image E of ±25.1 and image F of ±27.4.

factor referent to different observers. In turn, $B_{e,s}$ modelled as a Gaussian variable, where the mean $b_s$ corresponds to the bias of subject $s$, and the variance corresponds to the inconsistency ($v_s^2$) of that same subject.

$$X_{e,s} = x_e + B_{e,s},\ B \sim \mathcal{N}(b_s, v_s^2) \tag{3.4}$$

In order to estimate the unknown parameters $\theta = (\{x_e\}, \{b_s\}, \{v_s\})$, Maximum Likelihood Estimation (MLE) is used. MLE is a statistical method for estimating the parameters (*i.e*, mean and standard deviation) of a model, given some observations of the phenomenon in study [57]. This method relies on the maximization of a likelihood function. The likelihood function is a probability distribution, dependent on the variables that are being estimated, given the observations of the phenomenon in study [58]. As such, a log-likelihood function $L$, is defined in Equation 3.5. The parameters are estimated by solving $\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L$.

$$\begin{aligned} L &= \log P(\{x_{e,s}\}|\theta) \\ &= \log P(\{x_{e,s}\}|\{x_e\}, \{b_s\}, \{v_s\}) \end{aligned} \tag{3.5}$$

In light of the Belief Propagation algorithm [59] and assuming that the rating given by one observer is independent from the rating given by the other observers, Equation 3.5

can be rewritten according with Equation 3.6.

$$\begin{aligned} L &= \log \prod_{e,s} P(x_{e,s}|x_e, b_s, v_s) \\ &= \sum_{e,s} \log P(x_{e,s}|x_e, b_s, v_s) \end{aligned} \tag{3.6}$$

Now, since $X_{e,s}$ is a function of random normal variables, it can be itself written as a random normal variable, as stated in Equation 3.7.

$$X_{e,s} \sim \mathcal{N}(x_e + b_s, v_s^2) \tag{3.7}$$

Taking Equation 3.7 into account, the log-likelihood function $L$, can be rewritten in it analytical form, as shown in Equation 3.8.

$$\begin{aligned} L &= \sum_{e,s} \log \frac{1}{\sqrt{v_s^2 \cdot \sqrt{2 \cdot \pi}}} \cdot e^{-\frac{1}{2} \cdot \left( \frac{x_{e,s} - x_e - b_s}{\sqrt{v_s^2}} \right)^2} \\ &= \sum_{e,s} -\frac{1}{2} \cdot \log v_s^2 - \frac{1}{2} \log 2 \cdot \pi - \frac{1}{2} \cdot \frac{(x_{e,s} - x_e - b_s)^2}{v_s^2} \end{aligned} \tag{3.8}$$

In order to further simplify the log-likelihood function, it is possible to drop the constant part of Equation 3.8, as shown in Equation 3.9.

$$L \equiv \sum_{e,s} -\frac{1}{2} \cdot \log v_s^2 - \frac{1}{2} \cdot \frac{(x_{e,s} - x_e - b_s)^2}{v_s^2} \tag{3.9}$$

Then, this maximisation problem of the log-likelihood function expressed in Equation 3.9 is numerically solved by using the Newton-Raphson method [60]. This solution allows to estimate the true preference of the observer and to obtain a narrower confidence interval.

When applied to the subjective scores previously shown, the estimated preference of the observers says that in 84.3% of the cases, the observers choose the image rendered with the optimal patch size, no matter what is the image, with narrow confidence interval (at 95%) , that can be computed with the usual manner, assuming a normal distribution, as

shown in Equation 3.10.

$$x_e \pm 1.96 \cdot \frac{v_s}{\sqrt{S}} \tag{3.10}$$

Since the estimated preference values only present negligible variations (in the order of $10^{-4}$), the resultant confidence interval is very narrow, presenting a value of $28.6 \cdot 10^{-3}$.

In order to further analyse the relationship between the objective and subjective data and to evaluate the behaviour of the existing focus metrics, the Jaccard-Needham [61] and the Roger-Tanimoto [62] between the two scores will be computed. As such, the results shown in Table 3.1 have been binarised. In the case of subjective data, if the preference of the observers, for a given image, is higher than 50%, it is assigned value of 1 to that image. Otherwise, a value of 0 is assigned. Regarding the objective focus metrics, the variation $\Delta m$ is analysed. If $\Delta m$ is negative, the value 1 is assigned for a given metric applied to a given image. If $\Delta m$ happens to be positive, value 0 is assigned. Analogously, if the variation of the $q$, $\Delta q$, value is positive, number 1 is assign, otherwise, the value is 0.

With this binarised data, it is possible to compare how each metric behaves, against the subjective test scores, that are used as references. In order to numerically and systematically evaluate this performance, two measures of similarity of binary arrays where used: the Jaccard-Needham Index [61] and the Rogers-Tanimoto Index [62]. Note that, given the nature of the data under study, metrics such as the Russell-Rao [63], Sokal-Michener [64] and Dice [65] indexes yield the exact same results as the Jaccard Index. This evaluation of each focus metric is presented in Table 3.2. However, since the study carried out in [66] concluded that Rogers-Tanimoto index ($I_{RT}$) yields more accurate results than the other considered indices, all further discussion will be based on this indicator. As previously mentioned, $I_{RT}$ measures the similarity between two binary vectors, yielding 100% if both vectors are equal, and 0% if the vectors are complementary. It can be computed according to Equation 3.11, where $S_{ij},\ i,j \in \{0,1\}$ represent the number of occurrences of the pair $i,j$ in the corresponding position of both vectors.

$$I_{RT} = \frac{S_{11} \cdot S_{00}}{S_{00} + S_{11} + 2 \cdot S_{10} + 2 \cdot S_{01}} \tag{3.11}$$

From Table 3.2, it is possible to see that 20 out of 34 (59%) Roger-Tanimoto scores are 0% which means that the vast majority of the metrics totally contradict the subjective results. This tendency is even more pronounced in the metrics that are based upon the Laplacian of the image, where all of them present a score of 0%. The category that presents a higher percentage of non-zero scoring metrics, is the Statistics-based ones, reaching an average score of 39,5%. Further analysis shows that only 3 metrics attain a score above 50%:

| | | | Jaccard-Needham | Roger-Tanimoto |
|---|---|---|---|---|
| Subjective Tests | | | 100.00 | 100.00 |
| $\Delta q$ | | | 100.00 | 100.00 |
| Spatial Domain | Gradient | Gradient Energy | 0.00 | 0.00 |
| | | Gaussian Derivative | 33.33 | 20.00 |
| | | Thresholded Abs. Grad. | 0.00 | 0.00 |
| | | Squared Gradient | 0.00 | 0.00 |
| | | Tenengrad | 0.00 | 0.00 |
| | | Tenengrad Variance | 0.00 | 0.00 |
| | Laplacian | Energy of Laplacian | 0.00 | 0.00 |
| | | Modified Laplacian | 0.00 | 0.00 |
| | | Diagonal Laplacian | 0.00 | 0.00 |
| | | Variance of Laplacian | 0.00 | 0.00 |
| | Statistics | Chebyshev Moments | 33.33 | 20.00 |
| | | Absolute Central Mom. | 83.33 | 71.43 |
| | | Eigenvalues | 50.00 | 33.33 |
| | | Gray-level Variance | 66.67 | 50.00 |
| | | Gray-level Local Variance | 16.67 | 9.09 |
| | | Norm. Gray-level Local Var. | 66.67 | 50.00 |
| | | Modified Gray-level Var. | 0.00 | 0.00 |
| | | Histogram Entropy | 66.67 | 50.00 |
| | | Histogram Range | 83.33 | 71.43 |
| | Miscellaneous | Brenner's Measure | 0.00 | 0.00 |
| | | Image Contrast | 0.00 | 0.00 |
| | | Image Curvature | 50.00 | 33.33 |
| | | Helmli and Scherer's Mean | 0.00 | 0.00 |
| | | Steerable Filter | 33.33 | 20.00 |
| | | Spatial Frequency Measure | 0.00 | 0.00 |
| | | Vollath's Autocorrelation | 0.00 | 0.00 |
| | | Vollath's Standard Dev. | 83.33 | 71.43 |
| Transforms Domain | DCT | DCT Energy Ratio | 0.00 | 0.00 |
| | | DCT Reduced Energy Ratio | 0.00 | 0.00 |
| | | Modified DCT | 66.67 | 50.00 |
| | Wavelets | Sum of Wavelet Coeffs. | 0.00 | 0.00 |
| | | Var. of Wavelet Coeffs. | 0.00 | 0.00 |
| | | Ratio of Wavelet Coeffs. | 0.00 | 0.00 |
| | | Ratio of Curvelet Coeffs. | 33.33 | 20.00 |

Table 3.2: Evaluation of the Jaccard-Needham [61] and Rogers-Tanimoto [62] Indices, to compare the evaluations of the focus metrics with the subjective preferences.

Absolute Central Moments, Histogram Range and Vollath's Standard Deviation. For these 3 metrics, the maximum score obtained in these experiments is 71.4%. However, observing again Table 3.1, it is possible to conclude that Histogram Range presents very low values of $\Delta m$, with an average of 0%. This means that, effectively, Histogram Range does not differentiate between the image rendered with optimal patch sizes and the image rendered with the non-optimum patch sizes (one pixel larger). Although the same observation does not hold for Absolute Central Moments, with an average $\Delta m$ = -0.41% Vollath's Standard Deviation presents the best discriminative power, with an average $\Delta m$ = -0.81%, *i.e.,* almost 100% improvement over the former metric. Therefore, the obtained results lead to the conclusion that Vollath's Standard Deviation exhibits a better behaviour, as a focus metric than the others, when applied to images rendered from a focused plenoptic camera. Although it is not completely in accordance with the subjective results, it produces more consistent results, in-line with the expectations for a focus metric. It is also noticeable that the "quality factor", $q$, presents a Rogers-Tanimoto score of 100%, which is a good indicator of the image quality, in regard to the presence of sharp edges.

These results allow to conclude that applying standard focus metrics on extended depth-of-field images rendered from light field data might lead to erroneous assessments. In order to avoid the misleading behaviour of focus metrics due to artefacts generated at patch borders, a possible strategy is to apply these algorithms exclusively outside the affected regions. Additionally, a "quality factor", like the one defined by Equation 3.1, can provide a strength measure of such artefacts, which is useful to improve objective quality evaluation methods.

## 3.4 Final Remarks

In this chapter, 34 different focus metrics were considered and their performance was evaluated in the presence of artefacts generated by the rendering algorithm. It has been shown that most of the focus metrics are significantly affected by the high spatial frequencies artificially introduced by extended depth-of-field reconstruction, providing a wrong indication of better focus, due to overrating the focus level. This finding has been confirmed through a subjective evaluation process where the majority of observers preferred the images better focused with the optimal patch sizes, rather than those with higher objective focus level but lower perceptual quality.

Based on these results, a useful recommendation is to avoid the computation of focus level at the borders of micro-images, when traditional focus metrics are used in extended depth-of-field images. Furthermore, the smoothness of transitions between patches taken from adjacent micro-images should be considered in the evaluation methodology by specifically

measuring the gradients across patch boundaries.

# Chapter 4

# Method to Improve the Rendering Process

The rendering methods previously described yield unavoidable discontinuities across the patch borders of the rendered image, for one of two main reasons. Firstly, it does not take into account that lenslet images acquired with plenoptic cameras may exhibit regions where very low light intensity reaches the sensor, which corresponds to a lack of valid information in regions between micro-lenses. Secondly, the process to determine the optimal patch sizes cannot avoid the formation of noticeable transitions at patch borders after tiling. Overall, such patch-based effects contribute to degrade the performance of the rendering process used to obtain extended depth-of-field images by producing artefacts in the visual information. Therefore, in this chapter, both this problems were studied in depth and an algorithm to minimize each of the problems, thus improving the rendering process, was proposed.

## 4.1 Shortcomings of the rendering process

The quality of extended depth-of-field reconstructed images, obtained from the rendering process, suffers from its inherent limitations, which results in image artefacts, as explained in the following subsections.

### 4.1.1 Low light-intensity image regions

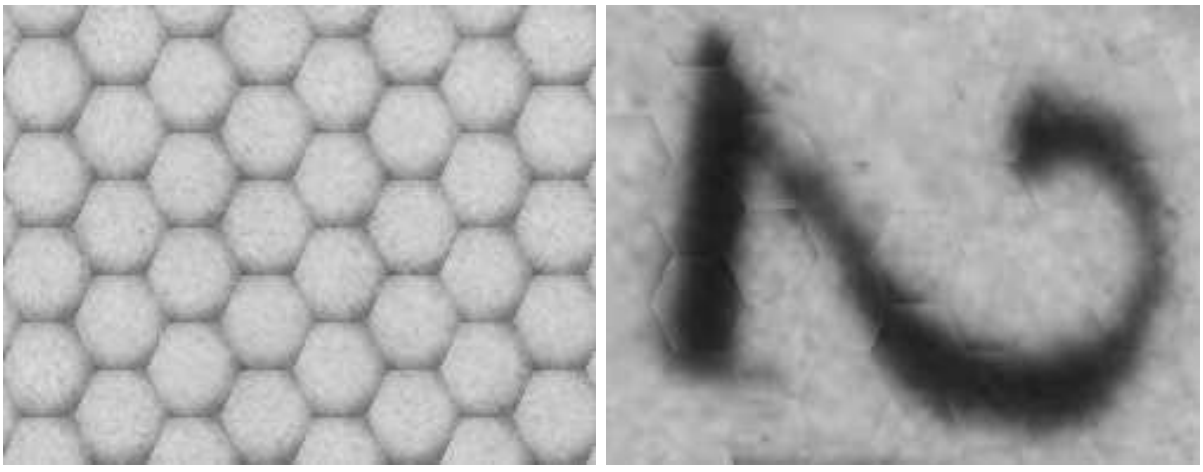
Lenslet images are generated by capturing the light ray intensity passing through the optical system of a plenoptic camera, comprised of a main lens and an MLA, placed

in front of the camera sensor [67]. Light rays incident on micro-lenses generate a micro-image, while those impinging on their borders or at interstitial space between micro-lenses are either highly attenuated or blocked, generating invalid information on the sensor. As a result, the lenslet images exhibit darker regions at the micro-images' borders.

In its minimization effort to solve Equation 2.8, the rendering algorithm may find patch dimensions that partially include some of the low light-intensity regions described above. This leads to patch-based artefacts. Figure 4.1a shows a rendered image of a uniform white surface. The expected rendered image would be a uniform white surface. However, what is seen in Figure 4.1a is a hexagonal pattern, corresponding to the micro-images. The dark boundaries of the micro-images are due to the vignetting effect of the micro-lenses.

### 4.1.2 Patch border artefacts

Although the optimum patch sizes found by the rendering algorithm described in Chapter 2 produce the smoothest possible transitions across patch borders, it is likely that some artificial sharp transitions appear along the matching borders. This effect, which can be observed in Figure 4.1b, has been addressed in [68] and [69], by proposing methods to overcome it, by using a depth map computed prior to rendering the extended depth-of-field image. However, this presents a chicken-and-egg problem [8], since a depth map is used to render the all-in-focus views, but one of the main goals of the view-rendering process is to convert the light field data to its 4D representation, in order to apply state-of-the-art algorithms for depth calculation. As such, a method for dealing with this unwanted transitions, with no need for compute a depth map form the lenslet data is proposed in the following sections.



(a) Effect of low light-intensity. (b) Patch borders artefacts.

Figure 4.1: Artefacts inherent to the rendering algorithm [8] (Image A).

## 4.2 Proposed Method to Improve the Rendering

The method proposed in this section to overcome the shortcomings described in Section 4.1 follows the pipeline presented in Figure 4.2. It is implemented as pre and post-processing, in regard to a modified patch-based rendering algorithm.

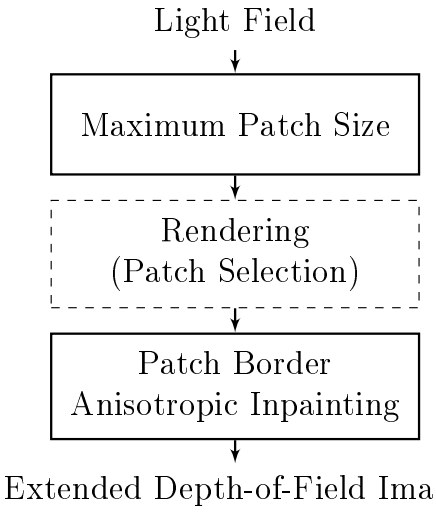



Figure 4.2: Proposed method including pre-processing (maximum patch size) and anisotropic inpainting of patch borders.

### 4.2.1 Finding the maximum patch size

In the pre-processing step, a maximum patch size is determined in order to exclude from the rendering stage those pixels of each micro-image border that belong to the low-light intensity regions described in Subsection 4.1.1. Taking into account the circular geometry of the micro-lens, this involves determining the circular region within the maximum micro-image radius that does not include the dark borders.

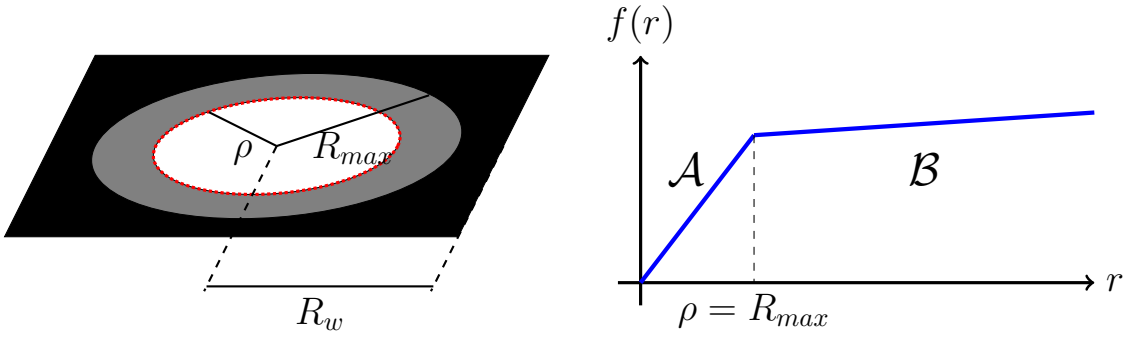
In order to estimate the maximum radial dimension, $R_{max}$, to be used in each micro-image $\omega(i,j)$, an auxiliary binary mask $K(\rho)$ (Equation 4.1) is defined such that an element $k \in K(\rho)$ is equal to 1 if its' radial distance $r_k$ is lower than an arbitrary radius $\rho$, otherwise $k = 0$, corresponding to mask elements located in the outer region of the circle of radius $\rho$, i.e. $0 < \rho < R_\omega$ and $R_\omega$ is the maximum possible radius, defined by the size of the micro-images (see Fig. 4.3a).

$$K(\rho) = \left\{ k \mid k = \begin{cases} 1, & \text{for } r_k \leq \rho \\ 0, & \text{for } r_k > \rho \end{cases} \right\} \tag{4.1}$$

To determine $R_{max}$, a function $f(\rho)$, defined for each micro-image is used. Such function is given by Equation 4.2, where $\langle \cdot, \cdot \rangle_F$ denotes the Frobenius inner product and $\omega$ is the set of all micro-images $(i, j)$ of a given lenslet image, following the notation used in Chapter 2.

$$f_{i,j}(\rho) = \langle \omega(i,j),\ K(\rho) \rangle_F \qquad \forall\, i, j \tag{4.2}$$

For each micro-image, Equation 4.2 is used to calculate the sum of all luminance values located inside the circle with radius $\rho$, thus it is monotonically increasing with respect to $\rho$ (region $\mathcal{A}$ in Fig. 4.3b). However, due to the low-light intensity occurring at the borders of each micro-image (see Figure 4.1a), the rate of change of $f(\rho)$ drastically decreases as $\rho$ increases towards the borders of each micro-image (region $\mathcal{B}$ in Figure 4.3b). This behaviour of $f(\rho)$, shown in Fig. 4.3b, is used to determine the largest radial dimension of the micro-image that should be used in the rendering process.




(a) Micro-image $w$.

(b) Plot of function $f(\rho)$.

Figure 4.3: Schematic showing the process for computing the maximum patch size for a micro-image (Equation 4.2).

Considering the properties of $f(\rho)$ discussed above, $R_{max}$ is determined as the value of $\rho$ that minimises the second derivative of $f(\rho)$, according with Equation 4.3.

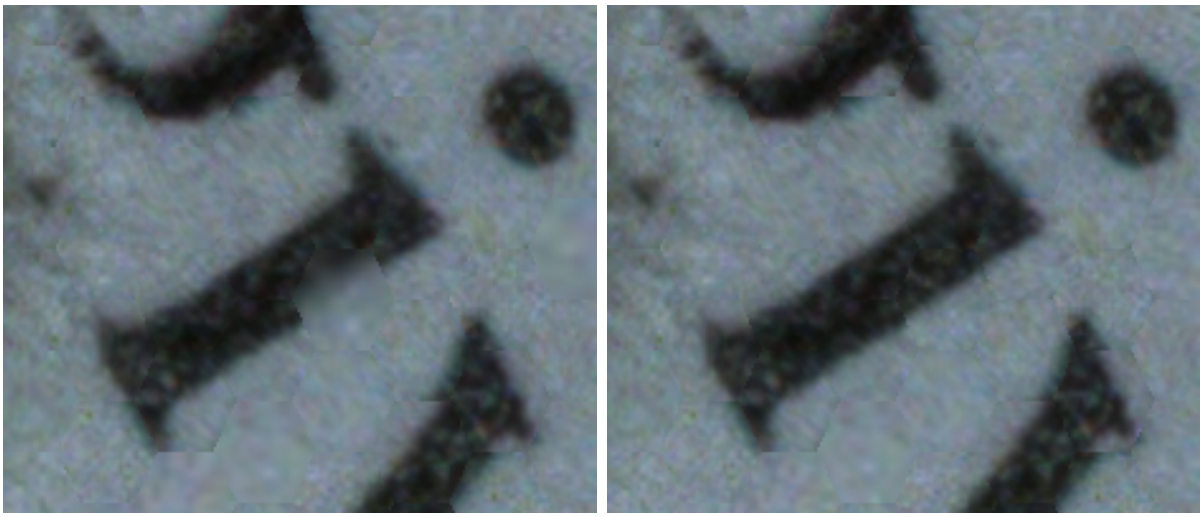
$$R_{max} = \underset{\rho < R_\omega}{\operatorname{argmin}} \frac{d^2 f(\rho)}{d\rho^2}. \tag{4.3}$$

### 4.2.2 Rendering — patch selection

In the original rendering algorithm [8], prior to determining the gradients of the Laplacian between a patch and its neighbours, all patches are downsampled to the size of the smallest possible patch. The border smoothness is then computed in this downsampled version of the refocused image, using Equation 2.7. Afterwards, the selected patches are upsampled

from their original resolution to the target resolution of the extended depth of field image. This is done to normalize the number of "source" pixels, before upsampling. Although, in principle, this is a useful reason, the ultimate goal of the algorithm is to render the best possible image quality. Therefore, it is worthwhile to compare the patches as they are going to be used in the final image, without any kind of normalization. Furthermore, as can be seen in Figure 4.4, without normalizing the patches yields better results.

In Figure 4.4, two images of the same scene and rendered from the same lenslet are shown. Figure 4.4a shows the result of rendering the view, when applying the discussed patch size normalization process, while in Figure 4.4b the view is rendered without any normalization. It is clear that the normalisation in Figure 4.4a gives rise to the choice of a smaller patch size in the edge of the letter "i", that seems out of place in the final rendered image. It does not happen in Figure 4.4b.



(a) (b)

Figure 4.4: Image 4.4a shows a view rendered with normalization of the patch size, and image 4.4b shows a view rendered without the normalization step.

A modified patch-based rendering method is proposed to improve rendering shown in Figure 4.2. The smoothest patch border transitions are determined using the original patch size, rather than a downsampled patch. In this case, less artefacts are produced because the patch used to maximize the borders smoothness is closer to the one used for tiling in the final rendering step.

### 4.2.3 Patch Border Anisotropic Inpainting

As previously described, after rendering the extended depth-of-field images some sharp transitions may still appear at the borders of the tiled patches (see Figure 4.1b). To minimise the effect of this type of artefacts, the final stage of the proposed method recon-

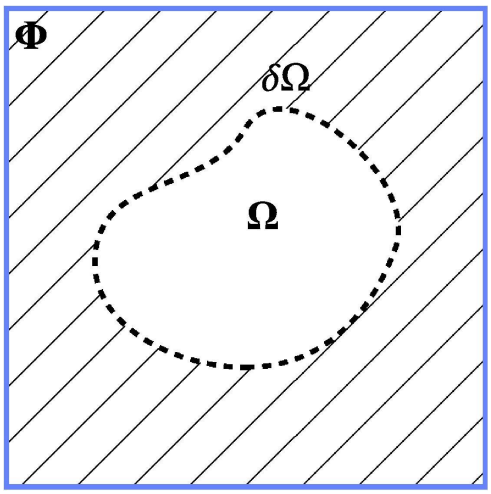



Figure 4.5: Generalisation of the inpainting problem: depiction of the know region $\phi$, the region to be inpainted $\Omega$ and boundary $\delta\Omega$ between the two (from [72]).

structs the corresponding affected regions by using a smoothness criterion, aiming to keep such border transitions as continuous as possible. This process only operates in limited regions defined as patch borders, since the remaining ones in the micro-image do not need any further correction. Therefore, the main objective of this stage is trying to propagate the accurate information available in the limited regions near the patch borders of the micro-images to the borders of the patches, such that the transitions between patches become continuous. This process is known as diffusion-based image inpainting [70]. One way to achieve this propagation is to use the heat equation [71], defined in Equation 4.4, where $u$ is the function encoding the spatial heat distribution in a given space, $\alpha$ is a positive constant that characterises the heat transfer rate, $t$ is the time variable and $\nabla^2(\cdot)$ is the Laplacian operator.

$$\frac{\partial u}{\partial t} = \alpha \nabla^2 u \tag{4.4}$$

Lets consider an image $I$ with a region to be inpainted $\Omega$ with boundary $\delta\Omega$ and a remaining region $\phi$, such that $\phi = I - \Omega$, as represented in Figure 4.5. Assuming that the contents of image $I$ can mimic the behaviour of the heat distribution function $u$ in Equation 4.4, and interpreting the region $\phi$ as fixed initial conditions (*i.e.,* the "heat" in that region is imposed to be constant through time), it is possible to propagate the information of region $\phi$ to fill region $\Omega$ by solving the heat equation, ensuring at the same time that the boundary $\delta\Omega$ is kept continuous.

Since it is not practical and/or efficient to solve Equation 4.4 algebraically, widely known numerical methods, such as the finite differences [73] can be used. In light of this method, the equation can be solve iteratively, by using Equation 4.5, where the time $t$ is modelled in terms of consecutive iteration steps, represented by $k$, and instead of heat distribution

$u$, the image $I$ is used instead.

$$I_{k+1} = I_k + \nabla^2 I_k \tag{4.5}$$

By definition, the heat reaches the equilibrium when the temperature of a given point does not change within a certain time interval. Thus, the stop criterium for Equation 4.5 is set, so that the equilibrium is reached when the absolute difference of $I$ between consecutive iteration is less than a threshold $\epsilon$, as shown in Equation 4.6

$$|I_{k+1} - I_k| \leqslant \epsilon \tag{4.6}$$

However, since the diffusion modelled by the heat equation is intrinsically isotropic, edges crossing through region $\omega$ are not taken into account and are thus not kept, making this region appearing somehow blurred, as shown in Figure 4.6b. In order to solve this problem, it is possible to modify Equation 4.4, in order to take the gradients of the original image into account, and thus preserve the edges. This kind of guided diffusion is called anisotropic diffusion [70]. In [74], it is suggested to modify the heat equation, according to Equation 4.7, where $\nabla$ denotes the differential operator and $\mathbf{D}$ a diffusion matrix.

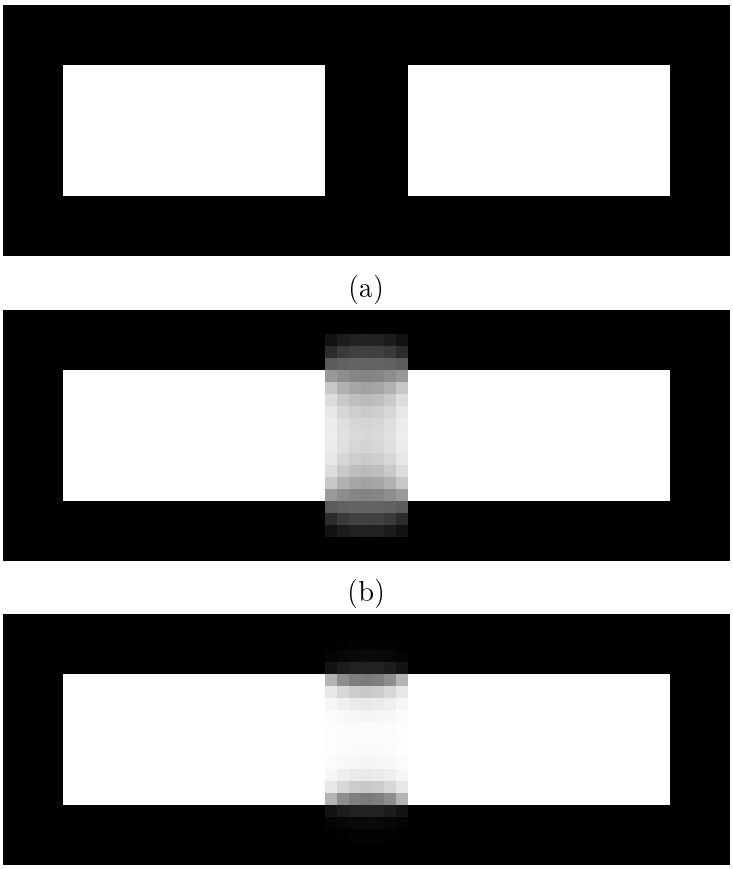
$$\frac{\partial I}{\partial t} = \nabla \cdot (\mathbf{D} \nabla I), \tag{4.7}$$

As with the heat equation, the finite differences method can also be applied to Equation 4.7, as proposed in Equation 4.8. This means that apart form the equation itself, the solution is attained by the same method and with the same stop criterium, since both equations are Partial Differential Equations (PDEs).

$$I_{k+1} = I_k + \nabla \cdot (\mathbf{D} \nabla I_k) \tag{4.8}$$

The diffusion matrix $\mathbf{D}$ is based upon an image descriptor, in this case the structure tensor, $T$. This can be computed pixel-wise, as represented in Equation 4.9, where $w$ is an averaging window function, and $I_x^\rho$ and $I_y^\rho$ are gradients of the Gaussian smoothed image $I$, with an inner scale $\rho$, in the $x$ or $y$ direction respectively. Usually, the operator $w^\sigma$ is a Gaussian kernel with an outer scale $\sigma$ [75].

$$T = \begin{bmatrix} w^\sigma * (I_x^\rho I_x^\rho) & w^\sigma * (I_x^\rho I_y^\rho) \\ w^\sigma * (I_x^\rho I_y^\rho) & w^\sigma * (I_y^\rho I_y^\rho) \end{bmatrix}, \tag{4.9}$$



(a)

(b)

(c)

Figure 4.6: Application of isotropic diffusion (Figure 4.6b) and anisotropic diffusion (Figure 4.6c) for inpainting the gap between the white bars on Image 4.6a.

The index $ij$ of the diffusion matrix $\mathbf{D}$ is determined from the eigenvectors $v_1 = [v_{11}, v_{12}]$ and $v_2 = [v_{21}, v_{22}]$ of $T$ and their associated eigenvalues $\lambda_1$ and $\lambda_2$ [76], as shown in Equation 4.10.

$$D_{ij} = \sum_{n=1}^{2} \lambda_n \cdot v_{ni} \cdot v_{nj}, \tag{4.10}$$

Figure 4.6 shows both the isotropic inpainting method (Figure 4.6b) and the anisotropic method (Figure 4.6c), applied to the gap between the two white bars on Figure 4.6a. In the case of the isotropic inpainting in Figure 4.6b, it is possible to see that the edge of the bars is not continuous. Instead, the black regions become white, giving the impression of blurriness. However, in 4.6c, there is a clear attempt to continue and connect the edge of both black bars, generating a sharper (and therefore much less linear) transition between black and white in the inpainted region. For this reason, the anisotropic inpainting method is chosen to fix the problem of the artificial edges that appear between neighbour micro-images.

In summary, the patch border regions (4 pixels wide, two pixel for each side of the border) are iteratively reconstructed, by propagating the information within the inner regions of the patches to its borders.

## 4.3 Results and Discussion

The performance of the proposed method was evaluated using the same set of six light field lenslet images, as in Chapter 3.

The effectiveness of the method proposed for smoothing the patch transitions of the extended depth-of-field images was numerically assessed in terms of an objective measure herein proposed. Additionally, visual observation was also considered for further validation. The validated quality metric $q$, defined in Equation 3.1, was used to numerically assess the reduction of the border artefacts, given that higher values of $q$ mean sharper border transitions and, consequently, lower rendering performance.

The objective measure $q$ obtained for extended depth-of-field images rendered with the proposed method was compared to that obtained for images rendered with reference algorithm [8]. The smoothness gain $G_q$, given by the ratio between $q$ of the reference algorithm and the proposed one, was used as performance metric. This ratio in shown in equation 4.11, where $I$ is the extended depth-of-field image rendered with the original

algorithm [8] and $J$ is the same image rendered with the proposed method.

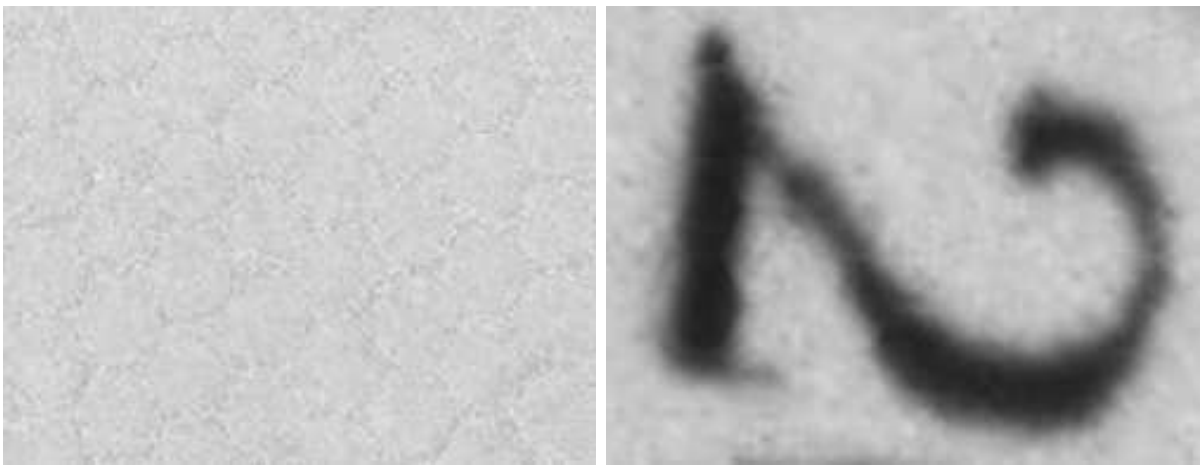
$$G_q = \frac{q(I)}{q(J)} \times 100 \tag{4.11}$$

The results, presented in Table 4.1 for images A to F, clearly show that the proposed anisotropic inpainting method drastically reduces the overall accumulated gradient over patch borders by many orders of magnitude. These results demonstrate that the proposed method efficiently smooths out the artefacts created by the commonly used rendering method.

| Image | **A** | **B** | **C** | **D** | **E** | **F** |
|---|---|---|---|---|---|---|
| $G_q$ | 111 | 98 | 624 | 986 | 350 | 106 |

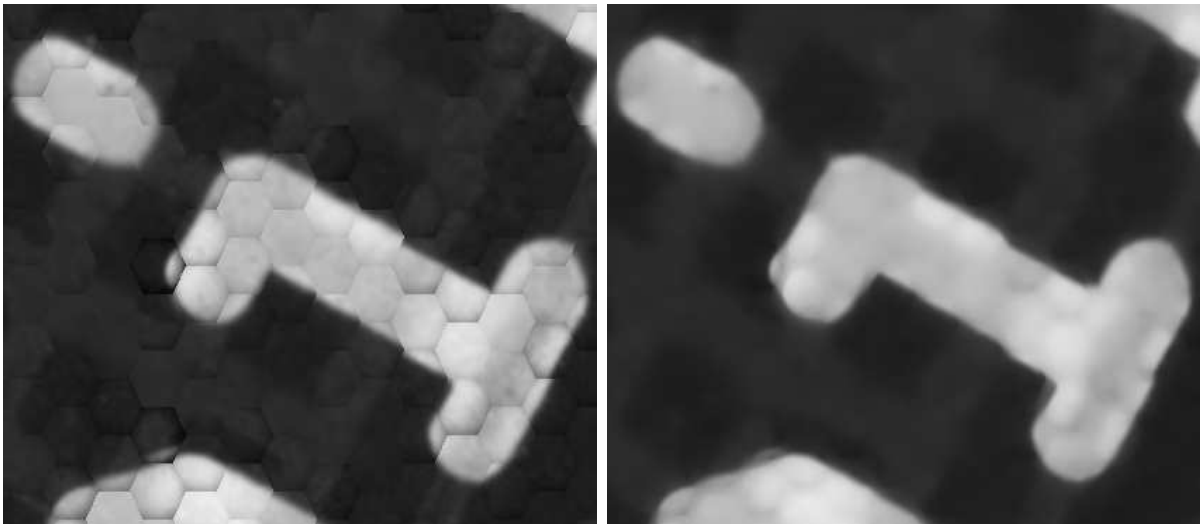
Table 4.1: Smoothness gain, $G_q$ of the proposed method for six images (A to F).

The visual effect of the images improved by the proposed method is shown in Figures 4.7, 4.8 and 4.9. From those results, the effectiveness of the anisotropic inpainting in reducing patch-based artefacts is quite obvious. The smoothing effect on patch borders can be clearly observed and its contribution for a better reconstructed image are quite evident. Similar results were achieved for all images used in the evaluation data set, as can be seen in Table 4.1.



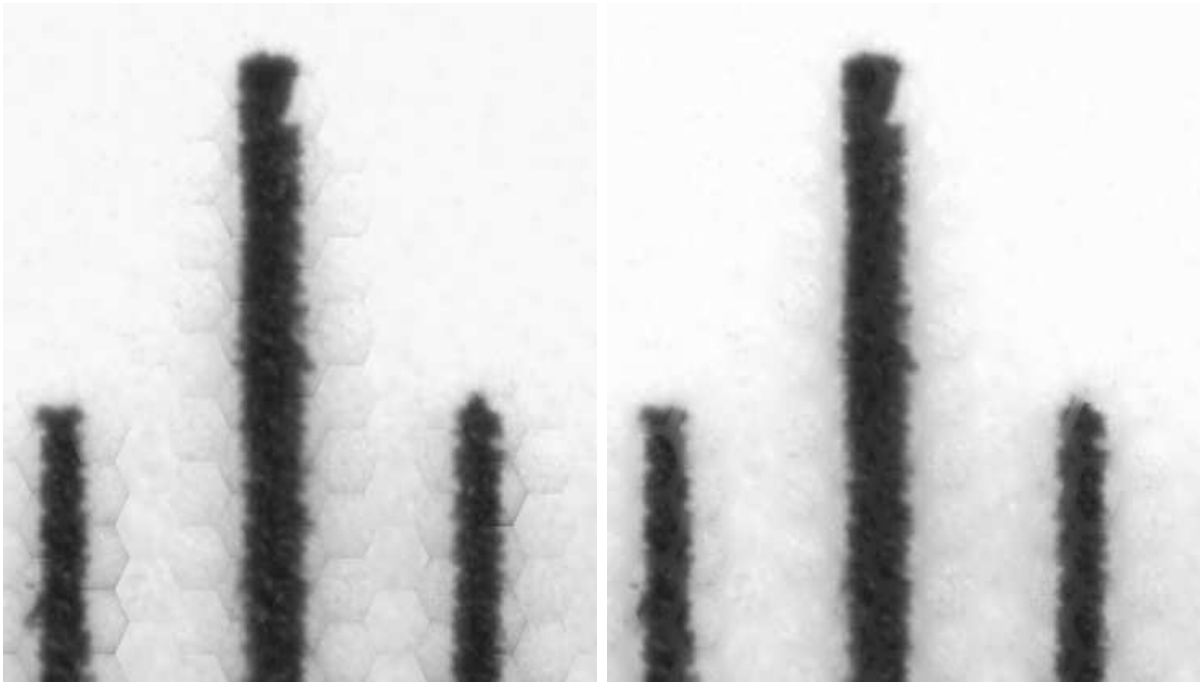
(a) Corrected effect of low light-intensity. (b) Mitigation of patch borders artefacts.

Figure 4.7: Images from Fig. 4.1 rendered with the proposed method (Image A).



(a) Rendering method in [8]. (b) Proposed method

Figure 4.8: Rendered image (Image B).



(a) Rendering method in [8]. (b) Proposed methods

Figure 4.9: Rendered image (Image F).

## 4.4 Final Remarks

The main shortcomings of patched-based algorithms for rendering extended depth-of-field were addressed in this chapter. In particular, an efficient solution, employing anisotropic inpainting, was devised that minimises the impact of low light-intensity regions and micro-image border gradients in order to improve extended depth-of-field image reconstruction. The proposed method, achieves better reconstructed images, significantly reducing the artefacts created by current patch-based methods. Overall, the quality and accuracy of refocusing algorithms may benefit from this method.

Furthermore, since the artificial edges have been mostly corrected by using methodologies described in this chapter, it is now expected that the application of current focus metrics to the rendered images yield correct results, without further adaptation of the metrics.

# Chapter 5

# Combining All-in-Focus Images from Extended Depth of Field Cameras

As previously explained in Chapter 2, focused plenoptic cameras are designed to achieve a better trade off between depth of field by and spatial resolution, in comparison with unfocused plenoptic cameras. In order to further improve the useful range of depth of field, *Raytrix* researchers introduced in [13] a focused plenoptic camera with extended depth of field.

In the case of the *Raytrix* R42, the improvements over the traditional focused plenoptic camera consists in the use of 3 types of micro-lens in the MLA, in order to increase the depth of field. Each type of micro-lens has a different focal length, *i.e.*, focusing at different depths. Therefore, by choosing the adequate lens type for each region of the captured scene, it is possible to effectively improve the depth of field, and yet obtain rendered images with high spatial detail.

## 5.1 Integrating the Lens Types into the Rendering

In practice, as proposed in [8], the three types of lens present in the lenslet image are treated independently, as if three focused plenoptic cameras with, MLAs with different focal lengths, were capturing the exact same scene. Hence, it is possible to render three similar images from a lenslet image captured with a focused plenoptic camera with extended depth of field. These images only differ on their focus plane and are slightly shifted apart in the horizontal direction, due to the fact that lenses of different types are not located in the exact same physical spot in the MLA (they are placed side-by-side, as shown in Figure 5.1).

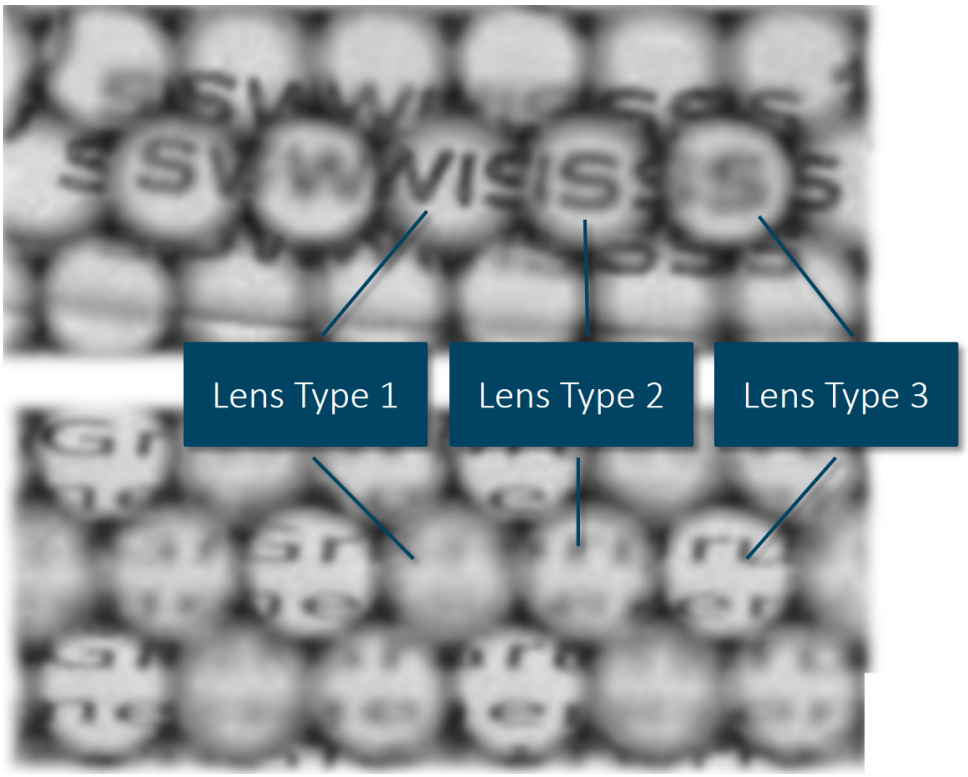



Figure 5.1: Representation of the lenslet structure typical of the focused plenoptic cameras with extended depth of field, where the 3 types of lenses with different focal lengths can be clearly seen[1].

In order to take advantage of the existence of three rendered images of the same scene, it is suggested in [8], that for rendering an extended depth-of-field image, only a correction of the horizontal offset between the images is necessary. Then the final rendered image is computed as the pixel-wise average of the three images. However, as the following discussion shows, when combining two or more images to improve the focus, it is always better to choose the better focused one, over a linear combination of the set of images.

In order to carry out such demonstration, lets consider an isotropic light source and a camera-like capturing scheme, with a main lens and a photo-sensor, as depicted in Figure 5.2. After going through the convergent main lens, the image is formed in-focus again, when both the dashed and the solid line converge. At this moment, the image should be an identical representation of the object. However, according to the model represented in Figure 5.2, the light only reaches the photo-sensor after the image is formed. The photo-sensor is represented in the figure by a vertical thick line. This means that the light, originally coming from a point source, is scattered through the region depicted in red (for now, lets assume that the scatter pattern follows a Gaussian shape, like depicted in the figure. This assumption will be explained later). In other words, the captured light is unfocused when it reaches the photo-sensor. Nonetheless, the light intensity emitted by the point light source, is the exact same that is received in the entirety of the red region

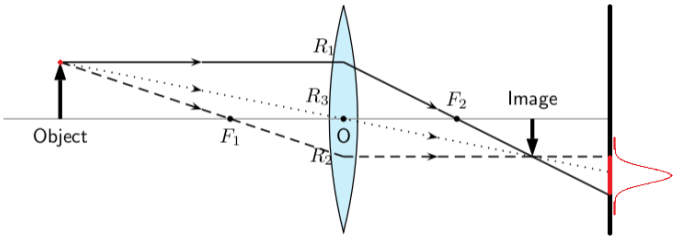



Figure 5.2: Unfocused camera-like acquisition system, simplified in 2-dimensions.

of the photo-sensor.

Thus, it is possible to integrate light intensity captured by the photo-sensor and it will always be less than the source light intensity, unless the integration interval covers light spread by the unfocused capturing system. If all the captured light is integrated, the attained value should be equal to the source light intensity. Although simplified in two dimensions, but without loss of meaning for the demonstration at hand, this inequality is shown in Equation 5.1, where $I_{source}$ represents the light intensity emitted by the light source, $I_{image}(x)$ represents the light intensity captured by the sensor at the vertical point $x$ and $c$ is the length of the region considered for integration.

$$I_{source} \geqslant \int_{-c}^{c} I_{image}(x)\,\mathrm{d}x \tag{5.1}$$

As stated in the previous paragraph, it is possible to conclude from Equation 5.1 that, if $c$ is big enough, strictly speaking, if $c = \infty$, the inequality becomes an equality, as expressed in Equation 5.2. This is the case, because all the light emitted by the source in eventually contained within the captured image, although it is more spread out.

$$I_{source} = \int_{-\infty}^{\infty} I_{image}(x)\,\mathrm{d}x \tag{5.2}$$

Now, according to [77], unfocused camera-like acquisition systems can be characterized by a Point Spread Function, which is basically the impulse response of the optical system. This means that the wider Point Spread Function, the more out of focus the optical system is. It is also mentioned in [77, 78], that Point Spread Functions can be approximately modelled by a Gaussian function. *Ergo*, Gaussian functions with higher standard deviation are able to model more out of focus systems. Thus, Equation 5.1 can be rewritten as shown in Equation 5.3, where $a = \sigma_a \cdot \sqrt{2\pi}$ and $\sigma_a$ is the standard deviation that parametrize a

[1]Image taken form `https://www.raytrix.de/wp-content/uploads/delightful-downloads/2015/07/Raytrix_LightFieldCamera_2015_print1.pdf`, as of $14^{th}$ of November 2018.

given optical setup with the corresponding degree of focus.

$$I_{source} \geqslant \int_{-c}^{c} \frac{1}{a} \cdot e^{-\left(\frac{x}{a}\right)^2} \mathrm{d}x \tag{5.3}$$

Considering now two different optical systems, characterized by $\sigma_a$ and $\sigma_b$, such that system with $\sigma_a$ is better focused than system with $\sigma_b$ (the Point Spread Function is narrower in the system with $\sigma_a$ than in the system with $\sigma_b$), which is mathematically equivalent to state that $\sigma_a < \sigma_b$, and thus $a < b$, given that $b = \sigma_b \cdot \sqrt{2\pi}$. It is then possible to establish the relationship between the light intensity, integrated for equivalent and equally sized regions, acquired by the two optical systems (Equation 5.4).

$$\int_{-c}^{c} \frac{1}{a} \cdot e^{-\left(\frac{x}{a}\right)^2} \mathrm{d}x \geqslant \int_{-c}^{c} \frac{1}{b} \cdot e^{-\left(\frac{x}{b}\right)^2} \mathrm{d}x \tag{5.4}$$

From Equation 5.4, it is clear that, for a small enough integration interval $c$, such that not all the light acquired by the optical system with $\sigma_b$ is considered, the left hand side of the above equation is always bigger than the right and side. Physically, this means that the light intensity captured by the system with $\sigma_a$ is more concentrated then in the system with $\sigma_c$. Thus, the system with $\sigma_a$ in closer to the punctual light source and thus is "more focused" then the system with $\sigma_b$.

In light of Equation 5.4, it is now possible to assess the validity of claiming that choosing the better focused image, is always better then considering the normalized linear combination of two unfocused images, regarding to the focus quality .In order to do so, Equation 5.4 was rewritten according to equation 5.5, where $\alpha, \beta \in \mathbb{R}$ are the weights of the linear combination.

$$\int_{-c}^{c} \frac{1}{a} \cdot e^{-\left(\frac{x}{a}\right)^2} \mathrm{d}x \geqslant \int_{-c}^{c} \frac{1}{\alpha+\beta} \cdot \left( \frac{\alpha}{a} \cdot e^{-\left(\frac{x}{a}\right)^2} + \frac{\beta}{b} \cdot e^{-\left(\frac{x}{b}\right)^2} \right) \mathrm{d}x \tag{5.5}$$

Solving Equation 5.5 yields the results that can be seen in Equation 5.6, where $\mathrm{erf}(\cdot)$ denotes the Gauss error function.

$$\mathrm{erf}\left(\frac{c}{a}\right) \geqslant \mathrm{erf}\left(\frac{c}{b}\right) \tag{5.6}$$

Given that $a < b$ as previously stated and, for the purpose of analysing Equation 5.6, lets consider $c = a$. The Equation then yields $\mathrm{erf}(1) \geqslant \mathrm{erf}(a/b)$, and since $a < b$, $a/b < 1$. Thus, considering the Gauss error function depicted in Figure 5.3, the equation is true for this particular case. In fact, given that $c < b$, and with help from Figure 5.3, it is possible

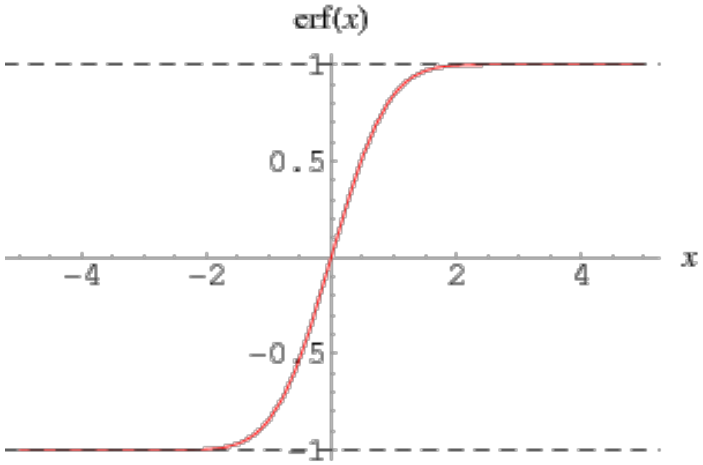



Figure 5.3: Plot of the Gauss error function.

to conclude that $\text{erf}(c/a) > \text{erf}(c/b)$ is always true, because $c/a > c/b$.

Thus, in light of the discussion about Equation 5.4, it is possible to conclude that choosing the "better focused" image between two unfocused images is always better, regarding focus quality, than combining the two images, as proposed in [8]. *Ergo*, in the remaining of this chapter, a method for choosing which type of lens is suitable for each region of the rendered image (*i. e.*, more in focus) is described.

## 5.2 Merging the Information from the 3 Types of Lens

In this Section, the workflow for joining the information os the three lens types into one extended depth-of-field image is described. This process is done in two steps: first, the three images generated from each type of lens need to be matched, since they present a horizontal offset between them, due to the spatial location difference of the lenses in the MLA. This type of image matching is commonly known in the literature as registration. Second, for each region of the captured scene, it is evaluated witch of the micro-lens' types is better focused. Then an image composed by these more in-focus parts of the original three rendered views is composed. This way, it is possible to avoid the linear combination approach suggested by [8], and take better advantage of the available data.

### 5.2.1 Image Registration

A first approach to find the offset between the images for correct registration, is to compute the cross-correlation between two of the images, and then find the maximum value of the cross correlation, which corresponds to the offset. Then, a simple horizontal shift of one

image would suffice to align both of them. However, this approach is limited to the resolution of one pixel, which might not be sufficient to properly align the images.

In order to have subpixel accuracy, it is possible to retain the same approach but upsample the image first. For example, to have half-pixel accuracy, the images would need to be upsampled to twice their original dimensions, to have quarter-pixel accuracy, the images would need to be upsampled four times their original dimensions, and so on. Therefore, this approach is tremendously memory consuming, in order to have relatively small increases in accuracy [79].

A much more efficient and fast algorithm is proposed in [79]. It defines the registration problem as a minimization of the normalized root-mean-square error (NRMSE), according with Equation 5.7, where $f(x, y)$ denotes the reference image, $g(x, y)$ the image to register, $(x_0, y_0)$ the offset between the two images, $^*$ the complex conjugate and $E$ the NRMSE.

$$\begin{aligned} E^2 &= \underset{x_0,y_0}{\operatorname{argmin}} \frac{\sum_{x,y} \left(g(x - x_0, y - y_0) - f(x, y)\right)^2}{\sum_{x,y} \left(f(x, y)\right)^2} \\ &= 1 - \frac{\underset{x_0,y_0}{\operatorname{argmax}} \sum_{x,y} \left(f(x, y) \cdot g^*(x - x_0, y - y_0)\right)^2}{\sum_{x,y} \left(f(x, y)\right)^2 \cdot \sum_{x,y} \left(g(x, y)\right)^2} \end{aligned} \tag{5.7}$$

Since the denominator of Equation 5.7 is constant for any two given images, only the numerator is relevant to the problem at hand. As such, the numerator can be rewritten according with Equation 5.8, where $F(u, v)$ and $G(u, v)$ represent the DFT of $f(x, y)$ and $g(x, y)$ respectively and $N$ and $M$ are the spatial dimensions of the image.

$$\underset{x_0,y_0}{\operatorname{argmax}} \sum_{u,v} \left(F(u, v) \cdot G^*(u, v) \cdot e^{i \cdot 2\pi \cdot \left(\frac{u \cdot x_0}{M} + \frac{v \cdot y_0}{N}\right)}\right)^2 \tag{5.8}$$

The algorithm proposed in [79] starts by computing the DFT of each image, after upsampling by a factor of 2. Then, Equation 5.8 is used to compute the offset between the images with an accuracy of half-pixel, in order to have a vague estimation of the final result. At this point, equivalent windows of 1,5 by 1,5 pixels, upsampled by an arbitrary desired factor of $\epsilon$, are compared between the two images, once again using Equation 5.8. The final step is to recover the two images, by preforming the inverse DFT on each of the images. This way, it is possible to compute the registration of two images, with a desired accuracy of $1/\epsilon$, in a efficient manner, memory-wise.

### 5.2.2 Multi-Focus Image Fusion

Image fusion methods are algorithms developed to merge several images of equivalent quality into one of higher quality, given that the definition of quality depends upon the characteristics of the images that are being improved [80]. More specifically, multi-focus image fusion techniques are devised to take the more in focus parts of two images of the same scene, and put them together, such that applying a focus metric to the final image yields a better result, then applying it to each of the original images individually, being the focus metric the definition of quality, in this case.

Therefore, in order to merge two images of the same scene, $f(x, y)$ and $g(x, y)$ with dimensions of $M$ by $N$, focused at different depth planes, the binary segmentation map $p(x, y)$ needs to be computed [81]. This map yield the value of 1 if the image $f(x, y)$ is to be chosen and the value 0 if $g(x, y)$ is the one to choose. This way, the resulting merged image $h(x, y)$ can be computed according with Equation 5.9, where $\cdot$ denotes pixel-wise multiplication.

$$h(x, y) = p(x, y) \cdot f(x, y) + (1 - p(x, y)) \cdot g(x, y) \tag{5.9}$$

In order to compute $p(x, y)$, it is proposed in [82] to first filter images according with Equation 5.10, where $*$ denotes the convolution operator and $|\cdot|$ the absolute value.

$$\begin{aligned} f_{filtered}(x, y) &= |H * f(x, y)| \\ g_{filtered}(x, y) &= |H * g(x, y)| \\ H &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{aligned} \tag{5.10}$$

The $H$ kernel in Equation 5.10 is design to emphasize the sharp edges present in the images. Now, the segmentation map can be computed, using a sliding window approach (in order to keep each maximization process feasible) of size $(2w+1)$ by $(2w+1)$, according with Equation 5.11, where $\Delta F = f_{filtered} - g_{filtered}$. $\Delta F$ allows to map the sharp edges of image $f(x, y)$ to the positive values and the edges of the image $g(x, y)$ to the negative values, allowing to condense the information of $f_{filtered}$ and $g_{filtered}$ in a single image.

$$W_{x,y} = \underset{W_{x,y}}{\operatorname{argmax}} \sum_{l=0}^{2w} \sum_{m=0}^{2w} \Delta F(x - w + l, y - w + m) W_{x,y}(l, m) \tag{5.11}$$

Equation 5.11 is then solved for each of the sliding windows, using the simplex method [83], subjected to the constraints transcribed in Equation 5.12, where $\epsilon$ is the spatial coherence threshold.

$$\begin{aligned} &|W_{x,y}(l,m) - W_{x,y}(l-1,m)| \leqslant \epsilon \\ &|W_{x,y}(l,m) - W_{x,y}(l,m-1)| \leqslant \epsilon \\ &0 \leqslant W_{x,y}(l,m) \leqslant 1 \end{aligned} \tag{5.12}$$

This constrains, force the variations of the segmentation map to be smooth. Hence, it effectively spreads the information contained in the sharp edges, to the regions where the original images are smooth. By joining all the sliding windows together, it is possible to Retrieve the whole segmentation map, according with Equation 5.13, where $N_w(x,y)$ is the number of windows that include the point $(x,y)$.

$$P(x,y) = \frac{\sum_{l=-w}^{l=w} \sum_{m=-w}^{m=w} W_{x+l,y+m}(w-l, w-m)}{N_w(x,y)} \tag{5.13}$$

The last step is to binarize the segmentation map, according with Equation 5.14.

$$p(x,y) = \begin{cases} 1, & \text{if } P(x,y) \geqslant 0.5 \\ 0, & \text{if } P(x,y) < 0.5 \end{cases} \tag{5.14}$$

After obtaining $p(x,y)$ from Equation 5.14 it is possible to use Equation 5.9 to merge the two images. However, in this application specifically, there is the need to merge three images. In order to do this, an intermediary image, which corresponds to the merge of two of the three images, needs to be computed. Now, to obtain the merge of the three images, the same algorithm is applied, but this time to merge this intermediary image and the third image.

## 5.3 Results

In order to evaluate the performance of the proposed method, a comparative metric, $G_F$ has been defined in Equation 5.15, where $f(\cdot)$ denotes a general focus metric, $I_0$, $I_1$ and $I_2$ denote the extended depth of field images rendered form lens types 0, 1 and 2 respectively, and $I_i \in \{I_0, I_1, I_2, I_{merged}\}$, where $I_{merged}$ denotes the merged image rendered form the

other three, using the method described in this chapter.

$$G_f = \frac{f(I_i)}{\max(\{f(I_0), f(I_1), f(I_2)\})} \cdot 100 \tag{5.15}$$

Analysing the application of Equation 5.15 for image A, for example, it is possible to quickly assess the better focused of the three original all focus images ($I_0$, $I_1$, and $I_2$), since it should yield the value 100, and how worse the others are focus wise, compared to it, since the other two shall yield values lower than 100. Analogously, if the value of Equation 5.15 computed for $I_i = I_{merged}$ is lower than 100, it means that it is worse focused than the better focused image of the set $\{I_0, I_1, I_2\}$, according with focus metric $f$. If the value is greater than 100, it means that $I_i = I_{merged}$ is better focused.

The measurements of focus, as described in the previous paragraph is shown in Table 5.1. In this case, the focus metric used was the Energy of Laplacian, since it is one of the more commonly used metrics. These measurements where carried out in the same way for the six images than where used in the previous chapters.

| Image | $I_0$ | $I_1$ | $I_2$ | $I_{merged}$ |
|---|---|---|---|---|
| **A** | 71 | 81 | 100 | 129 |
| **B** | 68 | 89 | 100 | 117 |
| **C** | 75 | 100 | 86 | 124 |
| **D** | 86 | 92 | 100 | 109 |
| **E** | 66 | 100 | 94 | 111 |
| **F** | 80 | 85 | 100 | 127 |

Table 5.1: Assessment of the focus improvement. Values in table computed using 5.15, with the Energy of Laplacian focus metric.

From Table 5.1, it is possible to conclude that the proposed method improves the overall focus for all the six images. In fact, it is possible to see that a minimum perceptual improvement of 9% in image D and a maximum of 29 % in image A. It is also possible to notice that, for the 6 test images, all present better focus for micro-lenses of type 1 or 2 than for type 0.

## 5.4 Final Remarks

In this chapter, the problem of combining the information of the three types of micro-lens was addressed. First, it was shown that any kind of normalized linear combination of two images is worse than picking the most focused image of the two, focus-wise. Upon this results, a method for selecting the most focused type of micro-lens for each region of

the scene was proposed, which raised the problem of image registration, which was also addressed. By using the combination of methods described in this chapter, an average improvement of 19,5 % in focus, measured by the Energy of Laplacian method.

# Chapter 6

# Conclusions

In this dissertation, rendering methods for focused plenoptic cameras with extended depth-of-field have been explored. First, in Chapter 3, the application of 34 different focus metrics on images rendered with focused plenoptic cameras was studied. It was then shown that the majority of focus metrics are inflated by the high spatial frequencies artificially introduced by the extended depth-of-field reconstruction algorithms. This results in a wrong indication of better focus, given that our study shows that images with more artefacts are deemed better focused then the ones with less artefacts. The objective results were confirmed by a subjective evaluation process, where it is shown that the majority of observers preferred the images better focused with the optimal patch sizes, over those with higher objective focus level but lower perceptual quality.

Then, in Chapter 4, the artificial high frequencies caused by the rendering algorithm in the region of transition between two neighbouring patches were mitigated, resorting to anisotropic image inpainting of these regions. Furthermore, a method for automatically calculating the maximum patch size for each micro-lens was proposed. This minimizes the impact of the low light-intensity regions characteristic of the rims of the micro-lenses in the rendered images. These improvements result in a significant improvement in the rendered images, both numerically and visually. Moreover, the focus measure problem presented in Chapter 3 is no longer an issue, since the effect of the artificial edges introduced by the rendering algorithm are no longer meaningful.

Finally, Chapter 5 describes a methodology for merging the three images of the same scene, that are the result of having three different types of micro-lens, in focused plenoptic cameras with extended depth of field. First, it was shown that any kind of normalized linear combination of two images is worse than picking the most focused image of the two, focus-wise. Then, an algorithm for sub-pixel image registration was described. This algorithm is based upon the search of the maximum of the correlation of the FFT of

two images. Lastly, an algorithm for merging the most focused regions of each the three images into one image was presented. This algorithm resulted in an average improvement of 19,5 % in focus, measured by the Energy of Laplacian method.

In summary, this thesis present a pipeline that improves the overall quality of rendered views from focused plenoptic cameras with extended depth of fields, both numerically and visually. Hence, since the view rendering process is the basis for other processing techniques, such as depth estimation, this pipeline provides a better foundation, than the original rendered algorithm, presented by [8].

# Bibliography


[1] M. Levoy, "Light field photography," Powerpoint, May 2013. [Online]. Available: https://graphics.stanford.edu/courses/cs178-13/lectures/lightfields-02may13.pdf

[2] H. Li, C. Guo, and S. Jia, "High-resolution light-field microscopy," in *Frontiers in Optics 2017*. Optical Society of America, 2017, p. FW6D.3. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=FiO-2017-FW6D.3

[3] J. Tian, Z. Murez, T. Cui, Z. Zhang, D. Kriegman, and R. Ramamoorthi, "Depth and image restoration from light field in a scattering medium," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2420–2429.

[4] K. Lynch, T. Fahringer, and B. Thurow, "Three-dimensional particle image velocimetry using a plenoptic camera," in *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*. American Institute of Aeronautics and Astronautics, Jan 2012.

[5] D. Weimer, H. Thamer, C. Fellmann, M. Lütjen, K.-D. Thoben, and B. Scholz-Reiter, "Towards 100micro components using plenoptic cameras," *Procedia CIRP*, vol. 17, pp. 847–852, 12 2014.

[6] H. Chen, M. A. Woodward, D. T. Burke, V. E. Jeganathan, H. Demirci, and V. Sick, "Human iris three-dimensional imaging at micron resolution by a micro-plenoptic camera," *Biomedical Optics Express*, vol. 8, p. 4514, 10 2017.

[7] A. Lumsdaine and T. Georgiev, "Full resolution lightfield rendering," Tech. Rep., 2008.

[8] S. Wanner, J. Fehr, and B. Jähne, "Generating epi representations of 4d light fields with a single lens focused plenoptic camera," in *Advances in Visual Computing*, G. B. et al., Ed. Berlin: Springer, 2011, pp. 90–101.

[9] "Integral photography," *Scientific American*, p. 164, aug 1911. [Online]. Available: http://www.tgeorgiev.net/Lippmann.pdf

[10] G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. Phys. Theor. Appl.*, vol. 7, no. 1, pp. 821–825, 1908. [Online]. Available: https://hal.archives-ouvertes.fr/jpa-00241406

[11] H. E. Ives, "Camera for making parallax panoramagrams," May 5 1936, uS Patent 2,039,648. [Online]. Available: https://www.google.ch/patents/US2039648

[12] A. Gershun, "The light field," *Journal of Mathematics and Physics*, vol. 18, no. 1-4, pp. 51–151, 1939. [Online]. Available: http://dx.doi.org/10.1002/sapm193918151

[13] C. Perwass and L. Wietzke, "Single lens 3d-camera with extended depth-of-field," pp. 829 108–829 108–15, 2012. [Online]. Available: http://dx.doi.org/10.1117/12.909882

[14] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, Feb. 1992. [Online]. Available: http://dx.doi.org/10.1109/34.121783

[15] R. Ng, "Digital light field photography," Ph.D. dissertation, Stanford, CA, USA, 2006, aAI3219345.

[16] M. , R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 924–934, jul 2006. [Online]. Available: http://doi.acm.org/10.1145/1141911.1141976

[17] D. E. Roberts, "History of lenticular and related autostereoscopic methods," *Parallax*, 2003. [Online]. Available: http://www.cs.toronto.edu/~kyros/courses/2530/papers/Lecture-02/Perwass2012.pdf

[18] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. MIT Press, 1991, pp. 3–20.

[19] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[20] D. Dansereau and L. Bruton, "Gradient-based depth estimation from 4d light fields," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, vol. 3, May 2004, pp. III–549–52 Vol.3.

[21] S. J. Koppal, *Lambertian Reflectance*. Boston, MA: Springer US, 2014, pp. 441–443. [Online]. Available: https://doi.org/10.1007/978-0-387-31439-6_534

[22] I. Ashdown, "Near-field photometry: A new approach," *Journal of the Illuminating Engineering Society*, vol. 22, no. 1, pp. 163–180, Winter 1993.

[23] HCI, "Light field structure," Website, Visited on: 20/05/2017, https://hci.iwr.uni-heidelberg.de/hci/softwares/light_field_analysis/structure.

[24] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolarplane image analysis: An approach to determining structure from motion," in *INTERN..1. COMPUTER VISION*, 1987, pp. 1–7.

[25] J. Bigun, "G.h.: Optimal orientation detection of linear symmetry," in *In: Proceedings of the IEEE First International Conference on Computer Vision, London, Great Britain*, 1987, pp. 433–438.

[26] B. Wilburn, "High-performance imaging using arrays of inexpensive cameras," Ph.D. dissertation, Stanford, CA, USA, 2005, aAI3153705.

[27] L. Stroebel, R. Zakia, I. Current, and J. Compton, *Basic Photographic Materials and Processes*. Focal Press, 2000. [Online]. Available: https://books.google.pt/books?id=b0nIELOwMs0C

[28] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera," Tech. Rep., Apr 2005. [Online]. Available: http://graphics.stanford.edu/papers/lfcamera/

[29] Q. W. Chunping Zhang, Zhe Ji, "Decoding and calibration method on focused plenoptic camera," *Computational Visual Media*, vol. 2, no. 1, p. 57, 2016. [Online]. Available: http://cvm.tsinghuajournals.com/EN/abstract/article_149266.shtml

[30] W.-S. Lai, "Introduction to the light-field camera," Powerpoint, Nov. 2013, http://disp.ee.ntu.edu.tw/class/20131129-LightField_intro.pdf.

[31] T. Georgiev and A. Lumsdaine, "Focused plenoptic camera and rendering." *J. Electronic Imaging*, vol. 19, no. 2, p. 021106, 2010. [Online]. Available: http://dblp.uni-trier.de/db/journals/jei/jei19.html#GeorgievL10

[32] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Journal of Optical Engineering*, vol. 32, pp. 2824–2836, 1993.

[33] J.-M. Geusebroek, F. Cornelissen, A. W. Smeulders, and H. Geerts, "Robust autofocusing in microscopy," *Cytometry*, vol. 39, no. 1, pp. 1–9, 2000.

[34] A. Santos, C. Ortiz de Solórzano, J. J. Vaquero, J. M. Peña, N. Malpica, and F. del Pozo, "Evaluation of autofocus functions in molecular cytogenetic analysis," *Journal of Microscopy*, vol. 188, no. 3, pp. 264–272, 1997.

[35] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, Dec 1995.

[36] E. Krotkov and J. P. Martin, "Range from focus," in *Proceedings. 1986 IEEE International Conference on Robotics and Automation*, vol. 3, Apr 1986, pp. 1093–1098.

[37] J. L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, and J. Fernandez-Valdivia, "Diatom autofocusing in brightfield microscopy: a comparative study," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 3, 2000, pp. 314–317 vol.3.

[38] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, Aug 1994.

[39] A. Thelen, S. Frey, S. Hirsch, and P. Hering, "Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation," *Trans. Img. Proc.*, vol. 18, no. 1, pp. 151–157, Jan 2009.

[40] M. V. Shirvaikar, "An optimal measure for camera focus and exposure," in *Thirty-Sixth Southeastern Symposium on System Theory, 2004. Proceedings of the*, 2004, pp. 472–475.

[41] P. T. Yap and P. Raveendran, "Image focus measure based on chebyshev moments," *IEEE Proceedings - Vision, Image and Signal Processing*, vol. 151, no. 2, pp. 128–136, April 2004.

[42] C.-Y. Wee and R. Paramesran, "Image sharpness measure using eigenvalues," in *2008 9th International Conference on Signal Processing*, Oct 2008, pp. 840–843.

[43] S. Pertuz, D. Puig, and M. García, "Analysis of focus measure operators in shape-from-focus," *Pattern Recognition*, vol. 46, 11 2012.

[44] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991.

[45] H. Nanda and R. Cutler, "Practical calibrations for a real-time digital omnidirectional camera," In Technical Sketches, Computer Vision and Pattern Recognition, Tech. Rep., 2001.

[46] F. S. Helmli and S. Scherer, "Adaptive shape from focus with an error estimation in light microscopy," in *ISPA 2001. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In conjunction with 23rd International Conference on Information Technology Interfaces (IEEE Cat.*, 2001, pp. 188–193.

[47] R. Minhas, A. A. Mohammed, Q. M. J. Wu, and M. A. Sid-Ahmed, "3D shape from focus and depth map computation using steerable filters," in *Image Analysis and*

*Recognition*, M. Kamel and A. Campilho, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 573–583.

[48] C.-H. Shen and H. H. Chen, "Robust focus measure for low-contrast images," in *2006 Digest of Technical Papers International Conference on Consumer Electronics*, Jan 2006, pp. 69–70.

[49] S. Y. Lee, J. T. Yoo, Y. Kumar, and S. W. Kim, "Reduced energy-ratio measure for robust autofocusing in digital camera," *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 133–136, Feb 2009.

[50] G. Yang and B. J. Nelson, "Wavelet-based autofocusing and unsupervised segmentation of microscopic images," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, vol. 3, Oct 2003, pp. 2143–2148 vol.3.

[51] H. Xie, W. Rong, and L. Sun, "Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2006, pp. 229–234.

[52] R. Minhas, A. A. Mohammed, and Q. J. Wu, "Shape from focus using fast discrete curvelet transform," *Pattern Recognition*, vol. 44, no. 4, pp. 839 – 853, 2011.

[53] Recommendation ITU-R BT.500-13: , "Methodology for the subjective assessment of the quality of television pictures," Geneva, Switzerland, Jan 2012.

[54] G. Howett and U. S. N. B. of Standards, *Size of Letters Required for Visibility as a Function of Viewing Distance and Observer Visual Acuity*, ser. NBS technical note, 1983. [Online]. Available: https://books.google.pt/books?id=dHTjrQEACAAJ

[55] Z. Li and C. G. Bampis, "Recover subjective quality scores from noisy measurements," in *2017 Data Compression Conference (DCC)*, April 2017, pp. 52–61.

[56] J. LI and P. L. CALLET, "Improving the discriminability of standard subjective quality assessment methods: a case study," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2018, pp. 1–3.

[57] J. Harris and H. Stocker, "Maximum likelihood method," *Handbook of Mathematics and Computational Science*, 01 1998.

[58] J. W. Harris and H. Stocker, *The Handbook of Mathematics and Computational Science*, 1st ed. Berlin, Heidelberg: Springer-Verlag, 1997.

[59] J. Pearl, "Reverend bayes on inference engines: A distributed hierarchical approach," in *Proceedings of the Second AAAI Conference on Artificial Intelligence*, ser. AAAI'82. AAAI Press, 1982, pp. 133–136. [Online]. Available: http://dl.acm.org/citation.cfm?id=2876686.2876719

[60] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*, ser. Texts in applied mathematics. Springer, 2002.

[61] P. Jaccard, "The distribution of the flora in the alpine zone.1," *New Phytologist*, vol. 11, no. 2, pp. 37–50. [Online]. Available: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1912.tb05611.x

[62] D. J. Rogers and T. T. Tanimoto, "A computer program for classifying plants," *Science*, vol. 132, no. 3434, pp. 1115–1118, 1960. [Online]. Available: http://science.sciencemag.org/content/132/3434/1115

[63] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, no. 2, pp. 159–203, 1948. [Online]. Available: http://www.jstor.org/stable/2983775

[64] R. Sokal, C. Michener, and U. of Kansas, ser. University of Kansas science bulletin. University of Kansas, 1958. [Online]. Available: https://books.google.pt/books?id=o1BlHAAACAAJ

[65] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945. [Online]. Available: http://www.jstor.org/stable/1932409

[66] B. Zhang and S. N. Srihari, "Properties of binary vector dissimilarity measures," 2002.

[67] M. Harris, "Focusing on everything," *IEEE Spectrum*, vol. 49, no. 5, pp. 44–50, May 2012.

[68] T. Georgiev and A. Lumsdaine, "Reducing plenoptic camera artifacts," *Computer Graphics Forum*, vol. 29, no. 6, pp. 1955–1968, 2010.

[69] P. Liu and R. Zhang, "Artifacts Reduction in Image Rendering of the Focused Plenoptic Camera," in *Proceedings of the 2015 International Conference on Intelligent Systems Research and Mechatronics Engineering*, vol. 121, no. Isrme, 2015, pp. 1292–1297.

[70] C. Guillemot and O. L. Meur, "Image inpainting : Overview and recent advances," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 127–144, Jan 2014.

[71] J. R. Cannon and F. E. Browder, *The One-Dimensional Heat Equation*, ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1984.

[72] M. A. Qureshi, M. Deriche, A. Beghdadi, and A. Amin, "A critical survey of state-of-the-art image inpainting quality assessment metrics," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 177 – 191, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320317301803

[73] C. Grossmann, H.-G. Roos, and M. Stynes, *Numerical treatment of partial differential equations. Revised translation of the 3rd German edition of 'Numerische Behandlung partieller Differentialgleichungen' by Martin Stynes*, 01 2007.

[74] Y.-L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 1539–1553, Nov. 1996.

[75] J. Weickert, *Anisotropic Diffusion in Image Processing*. Universität Kaiserslautern, 1996.

[76] D.-J. Kroon, C. H. Slump, and T. J. J. Maal, "Optimized anisotropic rotational invariant diffusion scheme on cone-beam ct," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 221–228.

[77] H. G. Dietz, "Out-of-focus point spread functions," pp. 9023 – 9023 – 11, 2014. [Online]. Available: https://doi.org/10.1117/12.2040490

[78] S. Stallinga and B. Rieger, "Accuracy of the gaussian point spread function model in 2d localization microscopy," *Opt. Express*, vol. 18, no. 24, pp. 24 461–24 476, Nov 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-24-24461

[79] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt. Lett.*, vol. 33, no. 2, pp. 156–158, Jan 2008. [Online]. Available: http://ol.osa.org/abstract.cfm?URI=ol-33-2-156

[80] H. Mitchell, *Image Fusion - Theories, Techniques and Applications*, 1st ed. Springer-Verlag Berlin Heidelberg, 2010.

[81] H. Li, B. S. Manjunath, and S. K. Mitra, "Multi-sensor image fusion using the wavelet transform," in *Proceedings of 1st International Conference on Image Processing*, vol. 1, Nov 1994, pp. 51–55 vol.1.

[82] A. Garnica-Carrillo, F. Calderon, and J. Flores, "Multi-focus image fusion by local optimization over sliding windows," *Signal, Image and Video Processing*, vol. 12, no. 5, pp. 869–876, Jul 2018. [Online]. Available: https://doi.org/10.1007/s11760-017-1229-x

[83] G. Dantzig, *Linear Programming and Extensions*, ser. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 2016. [Online]. Available: https://books.google.pt/books?id=hUWPDAAAQBAJ

# Appendix A

# Contributions

This appendix presents the contribution that resulted from the research work carried out in the scope of this dissertation.

- J. N. Filipe, L. M. N. Tavora, P. A. A. Assuncao, R. Fonseca-Pinto and S. M. M. de Faria, "Evaluation of Focus Metrics in Extended Depth-of-field Reconstruction," 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), Cagliari, 2018, pp. 1-6. doi: 10.1109/QoMEX.2018.8463366
- Sergio M. M. de Faria, Jose N. Filipe, Pedro A. A. Assuncao, Miguel O. Santos, Rui Fonseca-Pinto, Pedro M. M. Pereira, Luis M. N. Tavora, Felicidade Santiago, Victoria Dominguez, Martinha Henrique, "Light Field Image Dataset of Skin Lesions", 82nd ISO/IEC JTC1/SC29/WG1 (JPEG) Meeting, Lisbon, Jan 2019