MODELING AND APPLICATIONS OF THE FOCUS CUE IN CONVENTIONAL DIGITAL CAMERAS

Said David Pertuz Arroyo

Dipòsit Legal: T.1276-2013

Said Pertuz Arroyo

Modeling and applications of the focus cue in conventional digital cameras

PhD Thesis

Advisors:
Dr. Domènec Puig Valls
Dr. Miguel Ángel García García

Departament d'Enginyeria Informàtica i Matemàtiques

Universitat Rovira i Virgili
Tarragona
2013

UNIVERSITAT
ROVIRA I VIRGILI
Departament d'Enginyeria Informàtica
i Matemàtiques

Av. Paisos Catalans, 27
43007 Tarragona
Tel. 977 55 95 95
Fax. 977 55 95 97

FAIG CONSTAR que aquest treball, titullat "Modeling and applications of the focus cue in conventional digital cameras", que presenta Said Pertuz per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament d'Enginyeria Informàtica i Mathemàtiques d'aquesta universitat i que acompleix els requeriments per poder optar a Menció Europea.

Tarragona, 15 de juny de 2013

El director de la tesis doctoral

Dr. Domènec Puig Valls

A mis amados padres, Jorge y Nurys

Abstract

The focus of an image capturing system -either an artificial vision system, such as a photographic camera, or a natural vision system such as the human eye- plays a fundamental role in both the quality of the acquired images and the perception of the imaged scene. This thesis studies the *focus* cue in conventional cameras with focus control. This encompasses a wide range of acquisition devices, such as cellphone cameras, webcams, compact and single lens reflex digital cameras, surveillance cameras and the like.

A series of experiments is performed in order to understand and assess the different factors that affect the perception of focus by means of focus measure algorithms in computer vision. The aim of this task is to experimentally compare the advantages and limitations of different state-of-the-art approaches for the automatic estimation of the relative focus level in artificial vision systems.

After a deep review of the theoretical concepts behind focus in conventional cameras, it has been found that, despite its usefulness, the widely known *thin lens model* has several limitations for solving different focus-related problems in computer vision. In order to overcome these limitations, the *focus profile* model is introduced as an alternative to classic concepts, such as the near and far limits of the depth-of-field. The new focus model is based on an analysis of the image formation process by integrally incorporating concepts from wave optics and image digitization.

The new concepts and models introduced in this dissertation accurately describe the observed behavior in real systems and yield significant improvements with respect to previous existing approaches. These models are further exploited for solving diverse focus-related problems, such as efficient image capture (autofocus and focus sampling), depth estimation (through shape-from-focus), visual cue integration (through the novel shape-from-autofocus framework) and the generation of all-in-focus images (through focus stacking). The results obtained through an exhaustive experimental validation demonstrate the applicability of the proposed models for a wide variety of real and simulated scenarios.

Keywords: Focus measure, Autofocus, Focus stacking, Focus cue, Camera Calibration, Optics, Shape-from-focus, Depth-of-field, Defocus modeling.

Acknowledgements

Firstly, I would like to thank my advisors, Dr. Domènec Puig and Dr. Miguel Ángel García, not only for their guidance and fruitful discussions but also for their confidence in giving me the opportunity to carry on this thesis. Special thanks to Dr. Carme Julià for her advice while writing this dissertation and to Drs. Rodrigo Moreno (Linköping University) and Luis Pizarro (Imperial College London) for serving as external reviewers of this thesis.

Secondly, I would like the express my gratitude to Prof. Andrea Fusiello for giving me the opportunity to work at the Vision, Image Processing & Sound group (VIPS) during my stay at the University of Verona. I would also like to express my special gratitude to all, former and current members, of the Intelligent Robotics and Computer Vision (IRCV) group with whom I have shared these 5 years of work: Rodrigo, Marcela, Lin, Julian, James, Thomas, Xavi, Juan Manuel and Hatem; as well as the members of the VIPS group in Verona.

I am very grateful for the invaluable support received from innumerable friends at the URV and elsewhere. Special thanks to Rodolfo and Jaime for their expertise in photography. Some of the nice pictures in this thesis are owned by them. Special thanks to the Musoken Shotokan Karate club for all these years of practice. During all this time, Karate has been the balance that allowed me to keep my energy focused on my goals. I would also like to thank my "adoptive" family during my first two years in Tarragona, Isa, Beto and Isabel: feeling at *home* every time I came from work at the university has been priceless.

This thesis reached its completion with the financial support of the Universitat Rovira i Virgili (URV) through a pre-doctoral scholarship; the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) through scholarship 2010BE-DGR00; and Colciencias through scholarship 568.

Finally, I would like to express my deepest gratitude to my parents, Jorge and Nurys, my brothers Omar, Jorge and César, and my dear Yanine for their love, encouragement and support. Likewise, the support and affection from all my family, friends and people in Sincelejo and Colombia have been fundamental to me.

Contents

Abstract i
Acknowledgements ii
Contents iii

1 Introduction **1**
1.1 The focus cue 2
1.2 The focusing problem 5
1.3 Research directions 8
1.4 Objectives 9
1.5 Overview 10

2 Fundamentals **11**
2.1 Image formation 12
2.2 Focus-related tasks 22
2.3 Preliminary experiments 34

3 Focus measure **41**
3.1 Introduction 42
3.2 Focus measure operators 44
3.3 Magnification shift compensation 47
3.4 Comparative methodology 48
3.5 Experiments and discussion 54
3.6 Summary 63

4 Focus modeling **67**
4.1 Introduction . . . 68
4.2 New calibration and sampling methods . . . 70
4.3 Proposed focus profile model . . . 78
4.4 Predicting the behavior of focus . . . 86
4.5 Experiments and discussion . . . 88
4.6 Summary . . . 101

5 Confidence of the focus estimation **103**
5.1 Introduction . . . 104
5.2 Proposed reliability measure . . . 106
5.3 Efficient depth-map carving . . . 109
5.4 Improved focus stacking . . . 110
5.5 Experiments and discussion . . . 115
5.6 Summary . . . 126

6 Shape estimation from autofocus **129**
6.1 Introduction . . . 130
6.2 Focus signal model . . . 131
6.3 Shape from autofocus . . . 133
6.4 Experiments and discussion . . . 138
6.5 Summary . . . 141

7 Conclusions **143**
7.1 Summary of contributions . . . 144
7.2 Future research directions . . . 147
7.3 Publications . . . 149

Appendix A Focus measure operators **151**

Appendix B Defocus simulation **159**

Appendix C Focus profile: error sources **161**

References **163**

CHAPTER 1

Introduction

An important visual cue for both humans and artificial vision systems is *accommodation* or focusing. In human vision, accommodation is a monocular oculomotor mechanism that controls the configuration of each eye in order to keep the objects of interest sharply focused. The objects at different depths require different amounts of accommodation. Therefore, an accurate control of the degree of muscle strain in the eye is required in order to guarantee a satisfactory performance of the human visual system. Analogously to human vision, a wide range of artificial vision systems, including photographic cameras, microscopes, magnification glasses, etc., are concerned about the role of focusing during the image acquisition process for diverse focus-related applications.

This dissertation analyzes the focus phenomenon, its problems and applications in conventional cameras. In order to put the role of focus into perspective, this chapter starts with a brief review of different perceptual visual cues in section 1.1. The variables involved in the focus mechanism of conventional cameras are then discussed in section 1.2. Finally, the objectives of the thesis and an overview of the following chapters are presented in sections 1.4 and 1.5, respectively.

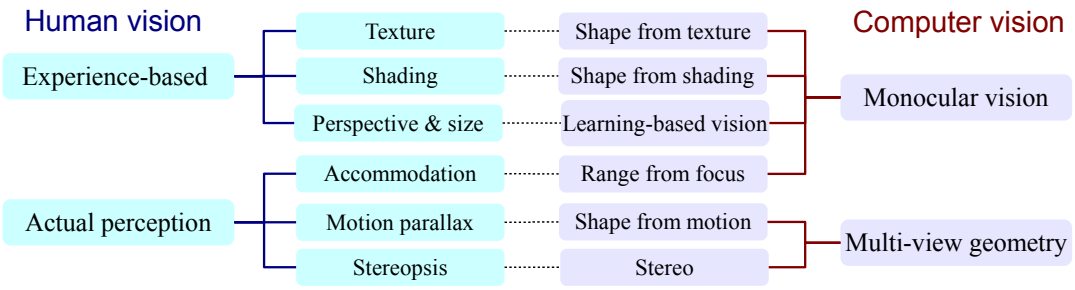


Figure 1.1: Basic visual depth cues in human vision and computer vision.

1.1 The focus cue

The visual system is fundamental for humans in order to perceive and understand the surrounding environment. At the lowest level of scene understanding, the task of depth perception, that is, determining the distance to the observed objects, is fundamental due to its key role in the interaction with the perceived world. In his seminal work on physiological optics, von Helmholtz (1924) distinguishes two main sources of depth information or *depth cues*: the first cue relies on experience and some familiarity with the nature of the perceived scene. Some mechanisms corresponding to this source are the determination of distance by means of the relative size of objects, their perspective, texture patterns and shading. The second depth cue involves an actual perception of depth, such as depth perception by motion parallax (e.g., by moving the head), stereopsis (binocular or stereo vision), and accommodation.

In computer vision, different visual cues have been exploited for depth and shape retrieval. From the optics perspective, the human eye is comparable to artificial vision systems in many different aspects (Navarro, 2009). As a result, many of the depth recovery techniques developed in artificial vision have been inspired by or show a close analogy with the human vision system (Tovee, 2008). For instance, Fig. 1.1 shows different depth recovery techniques applied in computer vision along with their analogous human depth cues. In this figure, depth recovery techniques in computer vision can be broadly grouped into two categories: those based on images captured with a single camera without changes in the viewpoint (monocular vision) and those based on multiple images captured from different viewpoints and/or different cameras. These techniques can be summarized as follows:

Shape from texture. When observing a 3D scene with a uniform texture, the imaged surface does not appear uniformly textured due to foreshortening and distance effects. The psychophysical experiments conducted by Gibson (1950) demonstrated that, indeed, the distance, shape and orientation of objects can

be inferred from the spatial properties of textured patterns on surfaces. This fact was subsequently exploited by Witkin (1981) for the recovery of shape and orientation from images of scenes with uniform texture patterns. Currently, this technique involves several steps in order to estimate local affine transformations of the imaged patterns and infer the surface orientation (Lobay and Forsyth, 2006).

Shape from shading. In painting, shading is the simplest technique for giving a sense of 3D in single 2D images. Horn (1975) exploited the fact that an image of a smooth object will exhibit gradations of reflected light intensity which can be used to determine its shape. In general, the problem of inferring shape from shading is underconstrained and requires modelling the image formation process based on assumptions on the illumination of the scene and the light reflectance properties of the surface or *albedo* (Zhang et al., 1999; Breuss et al., 2011). Recent approaches have tackled the problem of simultaneous estimation of albedo, shape and illumination from the shading of single images using priors (Barron and Malik, 2012).

Learning-based vision. The abstraction of shape, depth ordering, occlusions and object orientation from perceptual cues such as simple 2D line drawings and the relative size, deformation and aspect ratio of familiar objects is fundamental for understanding the 3D geometry of a scene. Humans are surprisingly effective at integrating various perceptive cues of this kind for successful scene interpretation and reasoning with relative ease. In contrast to previous and subsequent depth perception cues, learning-based perception is an open problem that involves a complex integration of multiple cues. As a result, there exist very different approaches depending on the exploited scene features, the integration strategies and the aim of the particular application. For an example of recent developments in this direction, refer to the work by Jia et al. (2012), Flint et al. (2011) and the references thereafter.

Range from focus. Accommodation is the mechanism by which the human eye guarantees that objects at a specific distance can be sharply imaged. Fixation is achieved by adjusting the ciliary muscles, which, in turn, change the power of the intraocular lens in order to see the fixated object more clearly. Therefore, by monitoring the degree of muscle strain in the eye, it is possible to determine the distance of the observed object. In computer vision, the principle of depth perception by accommodation (which is analogous to the autofocus mechanism of cameras) has been exploited by means of *range from focus* (chapter 2). In this case, the depth of the focused objects is determined by monitoring the current focus setting of the capturing device.

Shape from motion. When someone moves around, the retinal image of an observed 3D object is distorted since the object is successively seen from multiple directions. This depth cue was named the *kinetic effect* by Wallach and O'Connell

(1953). In computer vision, shape (structure) from motion infers the shape of the scene from multiple images captured from different viewpoints (Ullman, 1979). It is often assumed that the images correspond to continuous sequences acquired with a moving camera (or moving scene) and temporal coherence is exploited in the inference process. This technique, along with multiple view stereo, is widely known and has become a cornerstone in computer vision for the reconstruction of 3D scenes (Szeliski, 2011).

Multiple-view stereo. In his seminal paper, Wheatstone (1838) stated the fundamentals of stereopsis: in binocular vision, the perceived depth of an object is due to the disparity between the images projected on the two retinae. In general, multiple-view stereo can be exploited for the retrieval of 3D information with two or more images. This technique is tightly related to shape from motion with the exception that the camera pose is assumed to be known for each viewpoint (Seitz et al., 2006). This assumption can be relaxed by means of auto-calibration of the camera (Hartley and Zisserman, 2004), thus yielding a similar approach to shape from motion.

Albeit each depth cue can be regarded as an independent source of information, at a higher complexity level, the perception of depth and shape is a complex task that involves the combination and interaction of different cues. In the field of human vision, this fact was first reported by Wheatstone (1838). In order to effectively integrate different cues, it is critical to develop a deep understanding of each individual cue, its principles, advantages and limitations. In this sense, techniques based on multiple view geometry are arguably the most mature ones, whereas the accommodation (focus) cue has received relatively less attention. Thus, in many computer vision applications it is often assumed that either the images are sharply focused or the amount of defocus is negligible. However, at the most basic level, only an accurate knowledge and control of the focus mechanism of conventional cameras can guarantee the acquisition of high-quality images suitable for human analysis and computer vision tasks. Furthermore, in addition to accommodation (autofocus) and range from focus, focus is a a fundamental concept in several computer vision applications, such as *extended depth-of-field* and depth retrieval (*shape from focus* and *shape from defocus*). Therefore, focus is an important research field in computer vision.

In addition to conventional cameras, the focus mechanism has widely been studied for different devices, such as optical microscopes, telescopes and synthetic aperture radars, due to its critical role in the final quality of the acquired images. In this dissertation, the study of the focus phenomenon will be limited to digital cameras with focus control. This includes a wide range of devices, such as cell phone cameras, webcams, digital single-lens-reflex (DSLR) cameras, compact digital photographic cameras, surveillance cameras, and the like. The term *con-*

ventional cameras will be used in order to distinguish them from other specialized lens-camera systems, such as those used in microscopy imaging, telescopes, fish-eye lenses, catadioptric cameras, etc. In addition, the focus phenomenon will be addressed from an image processing-based perspective. Thus, technical issues such as the technological and manufacturing details are only discussed from their basic working principles, since we are particularly interested in the effects of focusing and defocusing on the acquired images and in the information that can be inferred from the images themselves.

1.2 The focusing problem

The behaviour of focus mainly depends on the system's optics (lens aperture, focal length and focus setting), the scene geometry (distance to the target) and the sensing device (sensor resolution, shutter speed and gain). The relationship and importance of these variables is a well known fact for most experienced photographers. In photography, the skills and capabilities of the photographer for manually adjusting these parameters determine the quality of the results in terms of focus. In contrast, in computer vision applications, and particularly in focus-related applications, it is critical to understand and achieve an accurate control of these variables in order to take maximum advantage of that visual cue.

A useful concept for describing the behaviour of focus is the *depth-of-field*. In the theoretical ideal case, when the focus of the camera is set to a certain distance, say u, only the objects at that exact position will be in perfect focus. Otherwise, the imaged object will undergo certain blurring due to defocus, which increases as the object departs from u. In practice, however, due to the limited resolution of real imaging systems, certain amount of blur can be neglected. As a result, all the objects within certain distance range, the depth-of-field (DOF), around the ideal focus position u, can be considered to be in focus. Let us now discuss the behaviour of the DOF as a function of four *focus parameters* commonly found in conventional cameras: the focus setting, the lens aperture, the focal length and the target position.

Focus setting. The most straightforward parameter that can be exploited in order to change the degree of blur of an imaged target is the *focus* of the camera itself. In particular, we are interested in cameras whose focus can be accurately controlled by means of a motorized mechanism. In these devices, focusing is achieved by changing the internal configuration of the lens-camera system. In the simplest case of single-lens cameras, this consists in changing the relative distance between the lens and the sensing device (e.g., the camera's CCD). In the most general case of a compound lens system (such as those found in any camera with zoom capabilities), this is achieved by changing the configuration of a set of fixed

Figure 1.2: Effect of camera's focus, u, on the captured image. Left: focus at $u = 0.7$ m. Right, focus at $u = 6.0$ m.

Figure 1.3: Effect of the camera's f-number, N, on the depth-of-field. Left: $N = 4.0$. Right: $N = 22$. A narrower aperture (right) yields a larger depth-of-field at the cost of a reduced illumination, which, in turn, is compensated for by means of the shutter speed and sensor's gain.

and moving lenses. In any case, the result is that the focus of the camera, u, moves from its current position. For instance, Fig. 1.2 shows the same scene captured with two different focus settings. It is clear that, depending on whether the camera focuses at the foreground or the background objects, the object of interest may be blurred or not.

Lens aperture. The diameter of the lens, or more exactly the effective lens aperture, is an important parameter for controlling the focus of a camera. As the aperture diameter is increased, the DOF decreases. From the perspective of geometric optics, this is due to the fact that, for a wider aperture, light rays that reach the camera must undergo a greater deviation in order to converge and form the image. As a result, the focused image rapidly diverges for positions away from the focus, yielding a shorter DOF (see Fig. 1.3). A side effect of changing the aperture is a change in the amount of incoming light. The effects of aperture on illumination are compensated for by means of the shutter speed (exposure time) and/or the gain of the electronic sensor (ISO). In photography, the aperture setting of the camera is often specified by the *f-number*, N, which is computed as the ratio between the lens focal length and the aperture diameter $N = f/D$. The bigger N, the narrower the aperture diameter.

Focal length. The lens focal length, f, determines the converging power of the lens. A large focal length yields a greater magnification (zoom). On the other hand, a large focal length implies that the incoming light rays passing through

Figure 1.4: Effect of the camera's focal length, f, on the depth-of-field. Left: target with $f = 70$ mm. Right: target with $f = 200$ mm. A larger focal length (right) yields a higher relative blur difference between the foreground and background objects.

Figure 1.5: Effect of target position, u_x, on the depth-of-field. Left: target at $u_x = 50$ cm. Right: target at $u_x = 70$ cm. The depth-of-field increases when the camera focuses distant targets (right).

the lens will undergo a large deviation before being focused due to the increased magnification. As a result, the objects away from the focus position are more rapidly defocused. For instance, Fig. 1.4 shows two images captured with different focal lengths. In both images, despite the distance between the foreground and background objects is the same, the relative amount of blur is different. In this figure, a change in perspective and magnification is also appreciable due to the difference in lens' power[1].

Target position. As previously stated, the change of focus in conventional cameras implies a change of the internal configuration of the lens-camera system. As a result, the DOF is different depending on the position of the focused target. For instance, the double arrow in Fig. 1.5 indicates the DOF limits. It is clear that, by changing the target position, not only the position of the in-focus position is changed but also the in-focus range (DOF).

[1]Note that, in photography, the depth-of-field is often not considered to be a function of the focal length, since the apparent reduced DOF is attributed to an increased magnification. Apart from this discussion, in this dissertation it is assumed that a change in the perceived blur (and hence the amount of focus that can be measured), corresponds to a change in the DOF.

1.3 Research directions

Within certain limits, the effect on the amount of defocus of the aperture size, the focal length, the focus setting itself and the target position can be controlled by changing the corresponding parameter or by moving the camera with respect to the target if possible. Notwithstanding, the perception of defocus also depends on less controllable variables such as the texture content, the illumination and contrast of the image, and the resolution of the system. As for the image content, it is clear that, in the extreme case of completely blank or light saturated targets, it is not possible to visually detect any change induced by defocus. This issue will be referred to as the *image content problem*. As for the resolution, one can think of a low resolution system in which the effect of defocus on the captured image can only be appreciated for large amounts of defocus. In addition, the exposure time and sensor gain have side effects on hand-shake blurring and image noise, respectively.

The challenge when solving focus-related problems in conventional cameras is to deal with the different variables involved in the focusing process in an integrated way. Although the effects of each of these variables are widely known and there exist theoretical models that describe their behavior, the development and application of practical and general models is still an open problem, as will be shown in the following chapters. For instance, the behavior of focus can be predicted by means of wave optics or ray-tracing software. Nevertheless, this requires an accurate knowledge of the imaging conditions and the exact geometry of the lens-camera system, which is an important restriction in most conventional cameras, since the internal lens design is often proprietary. There are simple approaches, such as the thin lens model, which can also be exploited in order to obtain a qualitative description of focus as a function of the lens focal length, focus setting and aperture. However, in the general case, even when there is full access to the focus controls of the camera, the real physical values of these parameters can only be known approximately at best. In most cases, their exact values and the internal lens design are unknown. As a result, conventional cameras usually behave as black boxes with unknown parameters.

An important concept that has been obviated in the previous discussion is *focus measure*. In order to assess and understand the effect of focus on an acquired image, it is necessary to be able to measure or estimate the degree of focus or blurring. In human vision, we intuitively detect the differences in sharpness of the images, independently of their content. In computer vision, this is achieved by means of algorithms that perform certain operations on the image content in order to compute the degree of focus.

In this dissertation, a new model for interpreting and predicting the effects of

the focus controls on the amount of blur in the images captured by conventional cameras is introduced. In particular, the classical thin lens model is extended in order to facilitate the calibration of cameras with unknown internal parameters. The new model is flexible and consistent, thus allowing a significant performance improvement of different focus-related applications such as autofocus, depth retrieval and image enhancement. As for the focus measurement, the working principle of different focus measure operators is analyzed and the factors affecting their performance are identified from a practical perspective.

1.4 Objectives

Bearing in mind the previous discussion about the focus controls and focus measure in computer vision, the goals of this thesis can be summarized in two main topics:

1. Assess the practical limitations and the factors that influence the detection of focus in conventional cameras. The aim is to identify the factors that affect the performance of focus measure algorithms according to their working principles.

2. Develop and validate a practical defocus model and its application to different tasks, such as extended autofocus and depth retrieval. The aim is to propose a new model for understanding the focus of conventional cameras in order to overcome some limitations of classical approches, such as the thin lens model.

Specific objectives

The specific objectives of this thesis are enumerated below:

1. Analysis of focus measure operators under different imaging conditions.

2. Development of an integral defocus model for conventional cameras.

3. Development of an efficient calibration method for defocus modeling and its application to autofocus and depth estimation.

4. Application of the developed defocus model for image enhancement through extended depth-of-field and improved depth estimation through depth-map carving.

1.5 Dissertation overview

The rest of this dissertation is organized as follows:

Chapter 2 introduces some fundamental concepts and notation as well as a review of the relevant literature. This chapter also provides preliminary experiments that illustrate common problems when performing focus-related tasks with conventional cameras.

Chapter 3 reviews and analyzes the performance of up-to-date algorithms used for computing the degree of focus of an image (or image pixel). The analysis is performed by taking into account different imaging factors, such as the image noise, contrast and saturation according to the working principles of the different algorithms.

Chapter 4 presents and validates the proposed defocus model for conventional cameras. The model is developed in order to provide a rich description of the effects of the focus controls, allowing a practical and robust calibration for those cases in which the parameters of the lens-camera system are unknown.

Chapter 5 validates the proposed defocus model through its application to different focus-related problems and applications, such as efficient focus sampling and extended depth-of-field. In addition, the concept of *reliability* is introduced in order to deal with the texture-content problem in focus measurement.

Based on the concepts introduced in previous chapters, Chapter 6 presents a new depth estimation approach, namely shape from autofocus, which allows performing basic scene understanding tasks such as shape estimation, depth ordering and object segmentation based on the focus cue of autofocusing cameras.

Finally, chapter 7 summarizes the contributions of this work and proposes future research directions and applications of the new concepts introduced in this thesis.

CHAPTER 2

Fundamentals

Since the very beginnings of photography with Daguerre's brand-new photographic process between 1835 and 1839, focus has been a concern and a key aspect for image acquisition. Beyond photography, focus has been studied in different fields, such as classical optics, computer graphics and computer vision. In optics, the main efforts have been devoted to obtaining an accurate understanding of the focus phenomenon in devices such as optical microscopes, laser beams and telescopes, in order to avoid defocus aberration. In computer graphics, the aim has been simulating its effects in order to produce realistic rendering of synthetic scenes. In computer vision, the first focus-related application has been autofocus, which is the analogous to accommodation in human vision, but has been extended to different tasks such as depth estimation and image enhancement.

This chapter reviews the image formation process and the models that allow understanding the defocus phenomenon in section 2.1. Section 2.2 summarizes the different focus-related applications in computer vision. Finally, in order to put the previous research into perspective, section 2.3 conducts some preliminary experiments, paying special attention to the different problems found in the acquisition of focus sequences and to the strategies for correcting them.

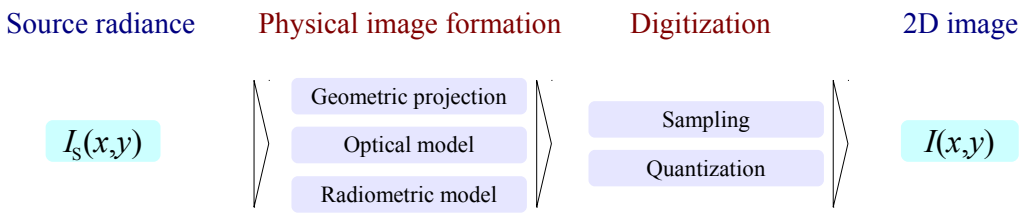


Figure 2.1: The image formation process.

2.1 Image formation

As shown in Fig. 2.1, the image formation process consists of three elements: the *3D scene* to be imaged (source radiance), the real image formed by an optical system (*physical image formation*), and the *digitization* that yields the final 2D digital image. By mathematically modeling this *imaging chain*, the characteristics and quality of the final image can be understood in terms of the different phenomena that take place during the image formation process (Fiete, 2010).

Source radiance. The image formation process in Fig 2.1 begins with the interaction between the light sources and the 3D scene. In order to be able to acquire an image of a real scene, a lens-camera system relies on capturing the energy of the light emitted, reflected and scattered by the objects (their *radiance*). The characterization and measurement of the light energy associated with some location or direction in space is known as *Radiometry*. For modeling purposes, the radiometric description of a 3D scene mostly depends on the scene illumination and its shape and radiometric properties (e.g., its reflectance). An accurate modeling and representation of the interaction of light at the scene level is one of the main challenges in computer graphics. In the scope of this dissertation, a real scene is simply regarded as a fixed 2D source radiance, $I_S(x, y)$.

Physical image formation. The electromagnetic energy of the light that reaches the camera's aperture is transformed by the optics of the system into a real image that is projected on the sensing device. Mathematically, this stage of the imaging chain can be sub-divided into three parts: geometric projection, optical model and radiometric model. The geometric projection describes the mapping from 3D scene points to 2D image points. In turn, the optical model describes the effects of diffraction and defocus of the system's optics on the formed image. Finally, the radiometric model describes how the energy of the formed image is distributed along the image plane.

Digitization. The last stage in Fig. 2.1, digitization, converts the real image that the system's optics projects on the sensing device, the image *irradiance*, in a discrete 2D representation. This stage involves two processes, namely sampling

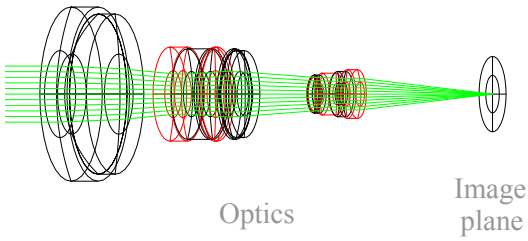


Figure 2.2: Ray tracing in a compound lens system. SLR-Zoom f = 72-145mm with optical compensation (Model WLZOOM-006 from Qioptiq's library)

and quantization.

In the following, the concepts related to the physical image formation and the digitization of the image are briefly revised. The discussion is driven towards those aspects which are relevant for understanding the focus mechanism of conventional cameras.

2.1.1 Optical model and radiometry

Most conventional camera-lens systems are complex in terms of the internal geometry and number of elements (Allen and Traintaphillidou, 2011). For instance, Fig. 2.2 shows the diagram of an SLR zoom lens. In this example, the system contains 10 single elements arranged in four different groups. Each lens group performs different optical functions such as zooming, focusing and compensation for optical artifacts. The response of these systems depends on the current internal configuration and varies along the image field. In order to obtain an accurate description of their behavior, lens designers rely on software packages, such as Code V or ZEMAX, in order to numerically perform physical-based ray tracing (Laikin, 2006). Although useful at the design stage, this approach is of limited application to conventional cameras from a user perspective, since the internal system geometry is often unknown. An analysis from the perspective of physical optics is required in order to obtain a more practical and meaningful understanding of the focus phenomenon.

According to wave optics, an optical system collects the light wavefronts and reshapes them as they are transmitted through each element of the optical system. For instance, in Fig. 2.3, the diverging wavefronts emitted from a distant point light source reach the optical system through the entrance pupil. The light waves leave the optical system through the exit pupil in converging wavefronts that form a focused image on the projection plane located at v. When the projection plane is displaced by a distance ϵ from v, a defocused image is formed. The plane

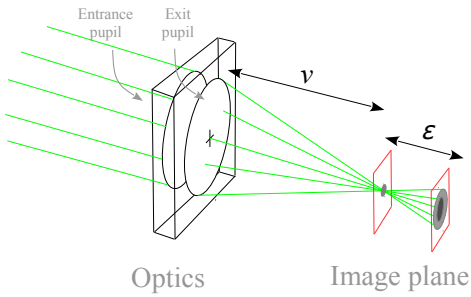


Figure 2.3: Schematic diagram of a defocused optical system. The incoming light from the left leaves the optical system in converging wavefronts that form an image on the image plane located at v. When the projection plane is displaced by a focus error ϵ, a defocused image is formed.

corresponding to the position with best focus, v, is often referred to as *focal plane* and depends on the characteristics of the particular optical system.

Being an electromagnetic wave, an accurate analysis of the propagation of light in optical systems requires and involved treatment of Maxwell's equations. However, it is possible to obtain meaningful closed-form solutions for some simple lens geometries, and by making some simplifications about the nature of light and the conditions of the imaging process. In the sequel, a bottom-to-top approach is followed, from the most specific and detailed cases to the simplest and most general ones. The implications and restrictions of the introduced simplifications are discussed. In the following discussion, scattering and absorption effects of optical elements are neglected.

With respect to the simplified model shown in Fig. 2.3, one could expect to find a perfectly focused spot when the position of the projection plane coincides with the focal plane at v. However, due to its wave nature, the light entering the optical systems departs from the ideal rectilinear trajectory as the incoming wavefronts are obstructed by the opaque pupil aperture. As a result, the image on the projection plane corresponds to a *diffraction pattern*, instead of on ideal infinitesimal blur spot.

A useful approach is to consider the case when the diffraction is shift-invariant under the *isoplanatic* assumption: when the target is fronto-parallel with respect to the ideal imaging system. In this case, based on the linearity of the wave propagation, the optics can be modeled as a linear shift-invariant (LSI) system, which can be characterized from its impulse response [1]. This allows describing the diffracted image, $I_D(x, y)$, as a simple convolution (Goodman, 1996; Fiete, 2010),

[1] Strictly, the shift-invariant assumption only holds for the paraxial geometrical focus plane (Gaskill, 1978). This paraxial geometrical optics will be discussed in subsequent paragraphs.

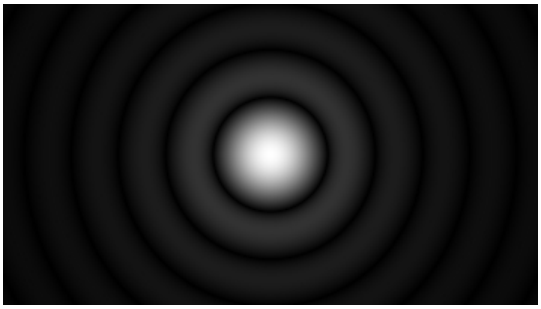


Figure 2.4: The Airy disc. The response of an ideal optical system to a monochromatic incoherent point light source is a diffraction pattern known as the Airy disc. The contrast has been equalized for displaying purposes.

$$I_D(x,y) = I_S(x,y) * h_0(x,y), \tag{2.1}$$

where $I_S(x,y)$ is the source radiance that originates the image and $h_0(x,y)$ corresponds to the impulse response of the LSI system. Naturally, for the 2D case, $h_0(x,y)$ is the response to a point light source and it is, therefore, referred to as the *point spread function* (PSF). The problem of finding the diffracted image from its source radiance now becomes finding the PSF of the system.

The diffraction on the focal plane is governed by the Fraunhofer diffraction, or far field diffraction. In this case, $h_0(x,y)$ for an ideal lens with circular aperture is a pattern known as the *Airy disc* (Ersoy, 2007; Goodman, 1996):

$$h_0(x,y) = \left(2\frac{J_1\left(\pi c_0 \sqrt{x^2+y^2}\right)}{\pi c_0 \sqrt{x^2+y^2}} \right)^2, \tag{2.2}$$

where $J_1(\cdot)$ is the Bessel function of the first kind and $c_0 = D/\lambda f$ is a constant that depends on the wavelength of light, λ, the aperture diameter, D, and the lens focal length, f.

The Airy disc is the PSF of the diffraction-limited optical system (Fig. 2.4). This occurs when the effects of diffraction dominate over all other aberrations. This is the case near the optical axis and when the image plane coincides with the focal plane. Notwithstanding, with respect to Fig. 2.3, we are also interested in the case when the *focus error*, ϵ, introduces a defocus aberration on the response of the system.

The study of a defocused optical system was pioneered by Hopkins (1955). The study of a defocused system requires an involved treatment of the Fraunhofer diffraction. For the sake of brevity, this analysis is not reproduced here since it can be found in the previous and subsequent references in this paragraph. However,

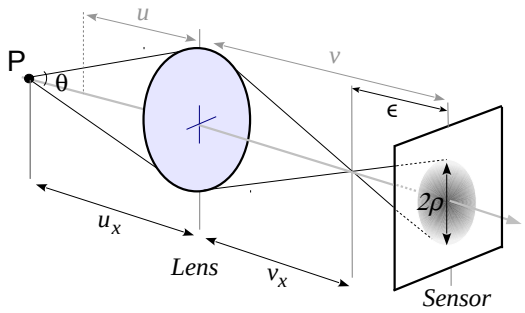


Figure 2.5: Geometrical optics approximation of defocus: the thin lens model.

an interesting remark of Hopkins' work is that a defocused system behaves as a low-pass filter whose cut-off frequency decreases as the focus defect increases. Subsequently, Brenner et al. (1983) and FitzGerrell et al. (1997) showed that a detailed mathematical analysis of an ideal defocused system can be achieved by means of the so-called Ambiguity function. Although the latter approach has shown to be quite useful, both theoretically and in practice for the design and development of range detection and extended depth-of-field systems (Dowski and Cathey, 1994, 1995), its extension to conventional cameras is not straightforward. Alternatively, further simplifications can be introduced in order to present a more practical (and intuitive) model suitable for its application to conventional cameras.

The electromagnetic field associated with visible light is characterized by its small wavelengths. Many optical problems can be simplified by neglecting the finiteness of the wavelength, yielding the so-called geometrical optics approximation (Born and Wolf, 1999). This simplification has three important implications (Menn, 2004): 1) light travels along straight trajectories (rays) and, therefore, diffraction effects can be neglected, 2) interference effects are not taken into account, and 3) ray tracing is invertible. With these assumptions, the defocused system in Fig. 2.3 can be analyzed in terms of simple geometric arguments, yielding the *thin lens* system shown in Fig. 2.5.

In Fig. 2.5, when a point light target, P, is located at a distance u_x from the camera, a perfectly focused image is formed at the image plane, located at v_x. When the angle θ is small, such that the first-order approximations $\sin(\theta) \approx \theta$ and $\cos(\theta) \approx 1$ can be used, we are dealing with the *paraxial approximation* of the geometrical optics (Born and Wolf, 1999). In this case, the relationship between the target position and the location of the focal plane follows the well known thin lens equation:

$$\frac{1}{f} = \frac{1}{u_x} + \frac{1}{v_x} \tag{2.3}$$

If the relationship in (2.3) does not hold, for instance by displacing the sensing

device in Fig. 2.5 by ϵ, the irradiance corresponding to P spreads over a blurring circle of radius ρ. This can be understood by the fact that, in this case, the position of the sensing device, v, does not coincide with the target's focal plane, $v_x \neq v$. As a result, the focus of the camera, u, and the target position, u_x, do not match. Naturally, defocus is corrected when $\epsilon \to 0$. It is possible to show that the blurring circle is given by (Subbarao, 1988; Pradeep and Rajagopalan, 2007):

$$\rho = v\frac{D}{2}\left(\frac{1}{f} - \frac{1}{u_x} - \frac{1}{v}\right) \tag{2.4}$$

The geometrical optics PSF of a defocused system is known as the *pillbox* function: a circular patch with constant irradiance (Horn, 1990; Subbarao and Surya, 1994; Born and Wolf, 1999):

$$h_\rho(x,y) = \begin{cases} \frac{1}{\pi\rho^2} & \text{if } \sqrt{x^2+y^2} < \rho \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$

Recall that (2.5) is the PSF of a defocused optical system with monochromatic illumination under the paraxial geometric approximation. In light of the central limit theorem, in order to take into account the joint effect of the different wavelengths in polychromatic incoherent illumination, as well as optic aberrations and lens imperfections, some researchers proposed to approximate the defocus PSF with a Gaussian instead of the pillbox function (Bass, 2010):

$$h_\rho(x,y) \approx \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right), \tag{2.6}$$

where σ is proportional to the blur radius ρ (Subbarao and Surya, 1994).

Notice that the thin lens model is derived from a first-order geometrical approximation of aberration-free systems. In practice, this model is quite useful and widely used to interpret the image formation process for many conventional imaging systems. Nevertheless, a third-order approximation, $\sin(\theta) \approx \theta - \frac{1}{3}\theta^3$ and $\cos(\theta) \approx 1 - \frac{1}{2}\theta^2$, is required in order to understand some common monochromatic aberrations known as third-order aberrations or Seidel aberrations, namely, spherical aberration, coma, astigmatism, radial distortion and field curvature (Born and Wolf, 1999). These aberrations change over the image field and mostly have an impact on the quality of the finally formed image by decreasing its sharpness. In particular, for focus-related applications such as shape-from-focus, the field curvature has been of special concern. In real systems, the shape of the focus plane is rather curved following a function known as the Petzval's curvature. When sensing this curved focus plane with a planar sensor, this causes fronto-planar objects to be radially defocused (section 2.3).

A final distortion effect not considered this far is radiometric distortion. Ideally, the irradiance of an image projected by an optical system should be uniform when imaging a perfectly uniform scene. Notwithstanding, the real irradiance has intensity variations due to radiometric distortion. The most common type of radiometric distortion is vignetting, which can be caused by occlusions of the lens aperture to the incoming light, displacements of the aperture from the front of the lens or variations of the off-axis illumination (Hasinoff, 2008). Depending on the application, it may be necessary to calibrate in order to compensate for the effects of vignetting (Kim and Pollefeys, 2008; Hasinoff, 2009). In this section, this effect is not presented in further detail since it is not directly related to focus.

2.1.2 Digitization

Once the optical system projects an image on the sensing device, the second element in the image formation chain is digitization. In simple terms, digitization consists in converting the energy of the projected image on a 2D matrix of values that represent the spatial irradiance. This stage can be sub-divided into two steps: sampling and quantization.

In the process of sampling, a continuous 2D irradiance field, $I_c(x_c, y_c)$, is mapped to a discrete 2D image, $I(x, y)$, whereas the sub-index c has been used to indicate continuity in the space domain (Voelz, 2010):

$$I_c(x_c, y_c) \rightarrow I(x, y) = I_c(mS_x, nS_y), \tag{2.7}$$

where S_x and S_y are the sample intervals, or sample spacing, in the x and y directions, respectively; $m = \{1, 2, \cdots, M\}$ and $n = \{1, 2, \cdots, N\}$ are the discrete image indices, assuming that the discrete 2D image has $M \times N$ samples (pixels).

According to the Nyquist-Shannon sampling theorem, the continuous domain irradiance, I_c, must be band-limited in order to avoid *aliasing* in the sampled image I. Aliasing is an undesired effect that can appear when representing a continuous domain function with a discrete number of samples. In order to avoid it, the following condition must be satisfied (Goodman, 1996):

$$\xi_c < \frac{1}{2S_x}, \ \eta_c < \frac{1}{2S_y}, \tag{2.8}$$

where ξ_c and η_c are the cut-off frequencies of I_c in the x and y directions, respectively.

Intuitively, (2.8) represents the fact that, when representing a continuous signal (the image irradiance) with a finite number of samples (the image pixels), the resulting digital image has a compulsory limited resolution.

The sampling rule in (2.7) is an ideal construction that takes samples of the continuous image irradiance on infinitesimal 2D spots at coordinates (mS_x, nS_y). However, in practice, the sensing device relies on processing a small area of finite size in order to estimate the energy of the irradiance corresponding to a single location. Specifically, each pixel consists of a photosensor aimed at measuring the image irradiance by integrating it both in the spatial and temporal domains. For illustration purposes, let us consider the simplest case of an individual imaging sensor consisting of a matrix of directly neighboring squared photoreceptors. In this case, the image irradiance at the projection plane is integrated over the area of each individual photoreceptor. Thus, the sample at coordinates (x, y) corresponds to (Jähne, 2004):

$$I(x,y) = G\beta\lambda\frac{\tau}{hc}\int\int_{\Omega(x_c,y_c)} I(x_c,y_c)\,dx_c\,dy_c, \tag{2.9}$$

where $\Omega(x_c, y_c)$ is a neighborhood that covers the $\Delta_x \times \Delta_y$ pixel area so that $\Omega(x_c, y_c) = \{(x, y)|x \in [x_c - \frac{\Delta_x}{2}, x_c + \frac{\Delta_x}{2}] \wedge y \in [y_c - \frac{\Delta_y}{2}, y_c + \frac{\Delta_y}{2}]\}$, G is the sensor gain constant, β is the quantum efficiency of the photosensor (the conversion rate between incoming photons and unit charges), λ is the wavelength of the incoming light, τ is the integration time (exposure) and h and c are the Plank's constant and the speed of light, respectively [2].

The effect of integrating over a finite spot in (2.9) can be interpreted as applying an averaging filter, whereas the image irradiance is low-pass filtered by the sampling spot. Thus, the sampling spot may be utilized to perform pre-sampling filtering for avoiding aliasing. As a result, the selection of the sampling steps, S_x and S_y, as well as the dimension of the sampling spot, $\Delta_x \times \Delta_y$ is an important design criterion that determines the final resolution of the imaging system (Pratt, 2007). It is important to remark that, as shown in the previous section, even in perfect focus, the imaged target is blurred due to the effect of diffraction. This diffraction blurring can be interpreted as a band-limiting effect of the system's optics. As a result, in order to optimize the acquisition process, the sensor resolution must be designed to be as close as possible to the resolution of the system's optics at the diffraction limit (Bass, 2010). This is an important fact that will be exploited in future chapters in order to propose a new defocus model.

During sampling, photodetectors work as transducers that convert the image irradiance into electric charge. In conventional cameras, the output of the detector is then amplified and converted to a discrete value. For instance, in 8-bit monochromatic images (gray-scale images), each pixel is assigned a value between

[2]Strictly, the quantum efficiency is a function of the wavelength: $\beta = f(\lambda)$. Therefore, in polychromatic illumination, (2.9) involves an integral as a function of λ (Jähne, 2004). For simplicity, this step has been omitted since it is not essential for the subsequent discussion.

0 and 255 according to the measured irradiance, being 255 the maximum allowed brightness and 0 the minimum one. Quantization is the process of scaling and discretizing the measured values. Both, photodetector measurement and quantization imply the addition of different noise types to the ideal sensed image. For instance, a CCD camera has several primary noise sources, such as *fixed pattern noise*, *dark current noise*, *shot noise*, *amplifier noise* and *quantization noise* (Healey and Kondepudy, 1994), which can be grouped into both irradiance-dependent and irradiance-independent sources. In that way, a noisy image I_n can be modeled as (Liu et al., 2008):

$$I_n = g(I + n_s + n_c) + n_q, \tag{2.10}$$

where I is the original image, $g(\cdot)$ is the camera response function (CRF), n_s is the irradiance-dependent noise component, n_c is the independent noise, and n_q is the additional quantization and amplification noise.

2.1.3 Geometric model

The basic geometry of most conventional cameras is a perspective projection that can be described by the well known pin-hole camera: an idealized camera whose imaging element consists of an infinitesimal small hole. According to this model, the world point $\mathbf{P} = [X, Y, Z]^T$ is projected to the point $\tilde{\mathbf{p}} = [\tilde{x}, \tilde{y}]^T$ at the image plane. It is clear that in this 3D to 2D mapping, one dimension is lost (the point depth Z). For simplicity, let us assume that the z-axis of the world reference system and the camera's reference system are aligned with the optical axis. The optical axis is an imaginary line that passes through the center of the image plane and the pin-hole. In matrix notation, the mapping $\mathbf{P} \rightarrow \tilde{\mathbf{p}}$ can then be expressed as (Hartley and Zisserman, 2004):

$$\tilde{\mathbf{p}} = \mathbf{KP},\ \mathbf{K} = \begin{pmatrix} m_x F & \alpha & x_0 \\ & m_y F & y_0 \\ & & 1 \end{pmatrix}, \tag{2.11}$$

where $\mathbf{K}$ is a referred to as the *intrinsic matrix* of the camera. The intrinsic matrix is unique for each camera and its parameters are the focal plane distance, F, scaling factors, m_x and m_y, the principal point, (x_0, y_0), and the skew parameter α that accounts for the non-orthogonality of the x- and y-axis of real systems (often the ideal case with $\alpha = 0$ is assumed). In the literature of multiple-view geometry, the products Fm_x and Fm_y are often referred to as *focal length* of the projection model. In order to avoid confusion with the lens focal length f, we have avoided this term here. In this dissertation, focal length refers to the lens focal length, f, unless otherwise is noted.

The pin-hole camera model governs the ideal perspective projection. Notwithstanding, real optical systems suffer from a number of inevitable distortions due to small imperfections in the manufacturing of the optic elements, optical aberrations and mis-alignments along the optical axis. As a result, a point is imaged at a larger distance (pin-cushion distortion) or shorter distance (barrel distortion) from the principal point (Tsai, 1987, 1986). This displacement is modeled as $[\delta x,\ \delta y]^T = (k_1 r^2 + k_2 r^4 + \cdots)\tilde{\mathbf{p}}$, where $r^2 = x^2 + y^2$ and $k_1,\ k_2, \cdots$ are the *radial distortion* coefficients. In addition to radial distortion, other types of distortion have also been considered such as tangential distortion (Heikkila and Silven, 1996), decentering and thin prism distortion (Weng et al., 1992). Thus, the last step of the image formation according to the pin-hole model corresponds to the mapping from the ideal undistorted image point $\tilde{\mathbf{p}}$ to the final image point $\mathbf{p} = [x, y]^T$.

The different parameters of the pin-hole model (intrinsic matrix and distortion coefficients), can be found by means of calibration. An early formulation of the current pin-hole model and *offline* calibration methods can be traced back to (Brown, 1971) and the references therein. Subsequently, a widely-known practical calibration procedure was proposed by Tsai (1987) exploiting calibration patterns of known dimensions in order to estimate the intrinsic and radial distortion parameters of the camera. A calibration procedure that includes tangential distortion parameters was proposed by Heikkila and Silven (1997). Flexible calibration procedures using planar calibration patterns were also proposed by Zhang (1999, 2000) and Sturm and Maybank (1999).

Alternatively to offline calibration, Fagueras et al. (1992) and Maybank and Faugeras (1992) pointed out the possibility of calibrating a camera based on the identification of matching points in several views of a scene taken by the same camera. This approach, known as *online* calibration, auto-calibration or self-calibration, is currently quite evolved and allows finding both extrinsic (camera orientation) and intrinsic parameters using images acquired with multiple cameras and multiple viewpoints (Hartley and Zisserman, 2004). Although self-calibration methods are more flexible than offline methods, lens distortion is usually neglected or assumed known (Remondino and Fraser, 2006).

Both, offline and auto-calibration methods consider that the camera has a fixed focus configuration. Nevertheless, for cameras with focus control, changing the focus setting yields changes in the internal parameters of the pin-hole model. This effect was first reported by Brown (1971) in photogrammetric applications when estimating the distortion parameters of objects at different distances from the camera. In his PhD thesis, Willson (1994) tackled the problem of modeling the perspective projection camera model of computer-controlled cameras, suggesting that each parameter of the pin-hole model changes as a function of the zoom-

focus setting. This implies that for each zoom-focus setting pair, there is a set of intrinsic parameters that characterize the perspective projection. In this case, the aim is to efficiently cover the zoom-focus space (Chen et al., 2000; Xian, 2006), and provide compact representations for the parameters of the model as a function of the camera setting (Sarkis et al., 2009).

Some researchers have reported that some intrinsic parameters, such as the principal point and distortion coefficients, can also be affected by changes in the lens aperture, in addition to the zoom and focus. As a result, some applications require an extensive parametrization in the zoom-focus-aperture space (Hasinoff et al., 2009). In contrast, some researchers suggest that the effects of changing the aperture setting of a camera is negligible (Li and Lavest, 1996; Chen et al., 2000; Sarkis et al., 2009). In practice and depending on the application, variations of the lens distortion with changes in focus can also be neglected, thus reducing the dimensionality of the problem (Fraser and Al-Ajlouni, 2006).

The perspective projection model is fundamental for accurately interpreting the images captured by a camera. In particular, in shape recovery applications, the pin-hole model allows translating the obtained shape into metric world units. Notwithstanding, in focus-related tasks, the effect of geometric projection can often be treated separately from the effect of focus. In the subsequent chapters, in order to center the discussions on the focus issue, it is assumed that the perspective projection model is either known or found by a previous offline calibration process, as described above. In addition, the variations of the pin-hole model parameters with changes in focus or aperture are neglected. The particular cases, requiring further considerations about the perspective projection model, are documented and discussed as they appear.

2.2 Focus-related tasks

Beyond the specific task of adjusting the focus of a camera in order to guarantee the quality of the captured image, this section reviews several relevant focus-related tasks and applications.

2.2.1 Focus measure

A critical stage of many focus-related problems is the estimation of the amount of focus, or *focus measure*, φ. As observed by Hopkins (1955), defocus can be interpreted as a low-pass filtering of the image radiance by the optical system. As a result, the energy of the defocused image $I(x, y)$ is a function of the amount of defocus of the system. In consequence, the computation of the focus measure value usually implies the application of a transformation to I, the so-called *focus measure*

operator (Subbarao and Tian, 1998). The energy of the transformed image over a region of interest is then used as an estimator of the focus level, φ:

$$\varphi = \int\int_{\Omega(x,y)} |I_t(x,y)|^2 \, dx \, dy, \tag{2.12}$$

where I_t is the transformed image after being processed by the focus measure operator and $\Omega(x,y)$ is the region of interest. Depending on the application, the region of interest can be as big as the whole image, as in autofocus applications, or as small as a neighborhood of 3×3 pixels for the computation of pixel-wise focus measures.

The transformation applied to the image is usually aimed at enhancing its spatial variations while being robust to noise. Focus measure operators tailored to this purpose exploit diverse concepts such as image derivatives, image statistics, the image Laplacian, wavelet transforms, the discrete cosine transform and the like. Due to its relevance for this dissertation, a detailed study of focus measure operators is presented in chapter 3.

An important limitation of focus measure operators is the so called *image content problem*. When the amount of texture in the image scene is too low, focus measure operators are unable to detect the change in focus level. In order to overcome this problem, the concept of *reliability* of a focus measure will be introduced in chapter 5. This concept is derived by taking into account that, even the best focus measure operator, will exhibit a non-ideal response to the variations of defocus due to inherent imperfections of the image acquisition process, such as optical artifacts and image noise. As a result, it is possible to define a reliability measure aimed at estimating how close the response of a focus measure operator is to the ideal noiseless aberration-free case. This reliability measure will allow an *a priori* prediction of when focus-dependent applications, such as shape-from-focus, are likely to work appropriately (chapter 5).

2.2.2 Autofocus

Due to the limited DOF, in the image of a scene captured through a finite aperture lens (either the lens of camera or a natural lens such as the human eye) only the objects within a certain distance range are sharply focused. Most digital cameras currently have an autofocus mechanism aimed at automatically adjusting the mechanical configuration of the lens-sensor system in order to capture sharp images without human intervention. In general, the aim is to reduce the time required for adjusting the parameters of the lens-sensor system in order to allow a fast and accurate capture. Autofocus has been an intensive research field in computer vision for many applications, including microscopy imaging (Wu, 2008), consumer photography (Yousefi et al., 2011; Jeon et al., 2010) and surveillance

(Yao et al., 2006). The different approaches for autofocus can be divided into three main methods: phase detection-based, active ranging-based and contrast detection-based.

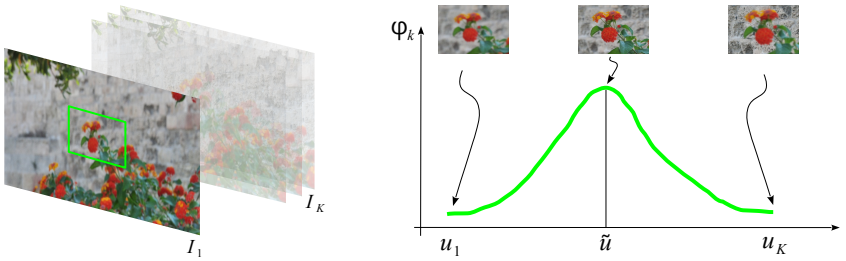
The *phase detection-based* method consists of splitting the image projected by the lens of the camera into two images that travel along optical paths of different length before being recorded by the camera's autofocusing circuitry (Bass, 2010; Kinba et al., 1997). This is achieved by means of a set of lenses and mirrors that work as light splitters. It can be shown that the split light beams correspond to spatially shifted images. This shift, in the spatial domain, is proportional to the difference between the current position of the lens and the position required for correct focus (the focus error). This focusing method is robust and fast, and is commonly found in professional photographic cameras such as single-lens reflex (SLR) cameras (Bass, 2010; Levoy, 2011).

The second method, the active ranging-based autofocus, exploits active range finders, such as infrared lights, in order to find out the distance between the target and the camera, so that the lens can be configured accordingly. This method is exploited by most modern compact digital photographic cameras and some webcams.

Finally, the contrast detection-based autofocus is arguably the oldest approach yet the most general and widely used. This method is aimed at inferring the correct in-focus distance only based on image processing. It is often used together with the other focusing methods whenever they fail to work as focusing mechanism (for instance, when illumination prevents active range finders to provide a reliable distance estimation or the light splitter cannot be activated in an SLR camera). It can be found in compact digital cameras, SLR cameras, surveillance cameras, microscopy imaging and webcams among others. In the sequel, only the contrast-based autofocus will be discussed.[3]

The problem of autofocus was first tackled by Horn (1968) and Tenenbaum (1971). Their approaches can be divided in two key steps: *focus measure* and *search strategy*. As previously stated, focus measure is aimed at processing the current image in order to measure its degree of focus. It is clear that, in order to be able to select the correct lens position, it is necessary to measure the degree of focus of the current image. The search strategy is a method for finding the lens configuration that maximizes the sharpness, or focus degree, of the acquired image. For instance, Horn (1968) proposed a global search strategy that consists in focusing the camera to a sequential set of positions that cover the whole focusing range: $\{u_k | k = 1, 2, \cdots, K\}$, where K is a predefined number of lens positions and $u_k < u_{k+1}$. For each in-focus position, a focus measure $\varphi(u_k)$ is then computed. Finally, the optimum lens position, $\tilde{u}$, is simply set to the position corresponding

[3]Contrast-based autofocus will be simply referred to as autofocus unless otherwise is indicated.



(a) (b)

Figure 2.6: Autofocusing. (a) Images captured at different focus settings with selected region of interest. (b) Focus measure values, φ_k, as a function of the focus setting, u_k. The aim is to maximize the focus measure value of the region of interest in order to capture the sharpest image.

the maximum focus measure.

$$\tilde{u} = \arg\max_{u_k} \varphi(u_k) \tag{2.13}$$

This autofocus approach is illustrated in Fig. 2.6. This figure shows a sequence of images captured at different in-focus positions (Fig. 2.6a) and plots the corresponding focus measure values (Fig. 2.6b). The set of images corresponding to different focus settings is known as the *image stack*. The global search method is an exhaustive, slow search of the maximum of the focus function. Notice from Fig. 2.6b that, ideally, the focus measure value, φ_k, is a smooth monotonic function of focus, u_k. Therefore different search strategies can be exploited for finding its peak without sweeping the whole focusing range. However, in practice, the focus function is corrupted due to image noise, inaccuracies in the measurement of focus, lens vibrations, etc, and the focus search is a non trivial task. In fact, most research regarding autofocus has been aimed at proposing either new robust focus measure operators or strategies for finding/inferring the maximum of the focus function. Proposed search strategies include global search, Fibonacci search (Krotkov and Martin, 1986; Xiong and Shafer, 1993), rule-based search (Kehtarnavaz and Oh, 2003; Tsai and Chen, 2012), hill-climbing (Ooi et al., 1990; He et al., 2003; Florea and Florea, 2011), fuzzy inference (Kuo and Chiu, 2011; Lee et al., 2008), successive interpolation (Geusebroek et al., 2000) and filter switching (Gamadia and Kehtarnavaz, 2012).

Notice that, in Fig. 2.6b, the focus function has been plotted as a continuous curve that is a function of the focus position. In practice, however, the focus function is only computed at discrete positions. The selection of those positions is fundamental for the success of the search strategy, as well as for the efficiency and speed of autofocusing. The process of selecting the focus positions is referred

to as *focus sampling*. In chapter 4, an efficient focus sampling procedure is introduced and applied for improving the speed of the autofocus process in conventional cameras.

2.2.3 Shape-from-focus

Notice that in autofocus, one can estimate the position of the target by finding the focus position at which the focus measure is maximized. This *range from focus* approach was pointed out by Jarvis (1983) and studied in more detail and implemented by Krotkov and Martin (1986) for performing range measurements of single objects. Later, Darrell and Wohn (1988) proposed to use image pyramids in order to compute coarse sharpness maps that assign a relative focus measure value to squared regions. Alternatively, instead of computing a focus measure in a large region, Nayar (1989) proposed the application of a pixel-wise focus measure in order to estimate the depth associated with each image pixel, thus allowing the computation of a complete depth-map. The latter approach is referred to as *shape-from-focus* (SFF).

From a historical perspective, the current SFF framework evolved in two stages. The first one corresponds to the SFF scheme introduced by Nayar (1989). In this scheme, the pixel-wise focus measure is computed using a small support window, or neighborhood, around the pixel in order to apply the focus measure operator. This approach works under the *isoplanatic* approximation that assumes that, within the pixel neighborhood, the scene shape can be considered as a fronto-parallel plane and, therefore, the focus measure is constant within the support window. In the second stage, in order to overcome the isoplanatic assumption, Subbarao and Choi (1995) introduced the concept of *focused image surface* (FIS) and used a 3D plane in order to approximate the shape of the object within the support window. Subsequently, Yun and Choi (1999) proposed a curved surface fit in order to approximate the FIS. Recent approaches have exploited different techniques in order to optimize the shape of the FIS such as neural networks (Asif and Choi, 2001), non-linear optimization (Ahmad and Choi, 2005; Mahmood and Choi, 2012), dynamic programming (Ahmad and Choi, 2007) or Bezier interpolation (Muhammad and Choi, 2010).

Algorithmically, the SFF problem can be divided into three main steps: focus stacking, peak detection and post-processing. Focus stacking consists in capturing an ordered sequence of images with different focus in order to generate an *image stack*, $I_k(x, y)$, for $k = 1, 2, \cdots, K$, where K is the total number of images. Similarly to autofocus, each captured frame is associated with a focus position, u_k. The process of changing focus and sequentially capturing an image at each focus position is referred to as *focus sweep*. Finally, the *focus stack*, or focus volume, is constructed by applying a pixel-wise focus measure operator to each frame of

the image stack, yielding $\{F_k(x,y)|k = 1, 2, \cdots, K\}$, where $F_k(x,y)$ is the focus measure associated with a pixel at coordinates (x,y) at the k-th frame of the image stack. The values of the focus measure for a pixel at particular coordinates (i,j) along all the focus stack are referred to as *focus function*, or focus measure vector: $\varphi_{i,j} = (F_1(i,j), F_2(i,j), \cdots, F_K(i,j))$. There is a focus function associated with each pixel of the captured scene.

The second step, peak detection, is aimed at finding the focus position corresponding to the maximum focus value for each focus function. Since the focus function is only defined at a discrete number of focus positions, peak detection is traditionally performed by fitting a model to $\varphi_{i,j}$ by means of interpolation. An initial depth-map, $z(x,y)$, is then constructed by assigning a depth estimation to each pixel. Thus, the depth corresponding to the pixel at coordinates (i,j) is estimated by simply finding the position of the maximum of the interpolated model:

$$z(i,j) = \arg\max_u \tilde{\varphi}_{i,j}, \tag{2.14}$$

where $\tilde{\varphi}_{i,j}$ is the function fitted to $\varphi_{i,j}$. The analytical model used to adjust the focus function is critical for the performance of SFF and still represents an open problem in the field. In fact, state-of-the-art models are empirical or have only been derived for microscopy imaging (Muhammad and Choi, 2012). Alternatively, in chapter 4, a new theoretical model based on the image formation process of a defocused image will be introduced and applied for improving the results of both shape-from-focus and extended depth-of-field (chapters 4 and 5).

In the final post-processing step, the depth-map is improved by means of filtering, smoothing or regularization techniques. For this purpose, in addition to the FIS optimization-based approaches referenced above, some authors have proposed median filtering (Nayar and Nakagawa, 1994), region-based and bilateral filtering (Shoji et al., 2006; Aydin and Akgul, 2008), regularization using Markov random fields (Sahay and Rajagopalan, 2008; Gaganov and Ignatenko, 2009), regularization incorporating defocus information (Pradeep and Rajagopalan, 2007) and similar strategies (Mahmood et al., 2008; Shim et al., 2009; Shim and Choi, 2010).

2.2.4 Focus stacking

The limited DOF of optical systems is often a problem in image acquisition since it leads to defocusing of those parts of the depicted scene that are not comprised within the in-focus limits. As previously illustrated in Fig. 1.2, when the focus of the camera changes, objects at different depths are selectively brought in and out of focus depending on their distance to the current in-focus position. Since the limited DOF is a common problem for most optical imaging systems, extending the depth-

(a) (b) (c)

Figure 2.7: Focus stacking. (a) 3rd frame of a focus stack. (b) 13th frame of a focus stack. (c) All-in-focus image generated by focus stacking. ©HTTin, 2009.

of-field has received much attention in the research community. Early solutions consisted in modified acquisition devices tailored to the acquisition of large DOF images. For example, McLachlan (1964) proposed a special illumination system in order to capture extended DOF images in microscopy. Currently, state-of-the-art solutions include modified cameras and modified optic systems that allow extending the DOF by non-conventional capture processes (section 2.2.7).

In this dissertation, we are specifically interested in the solution of this problem with conventional cameras. In the next chapters, the applications and discussions refer to extended depth-of-field in conventional cameras unless otherwise is noted. In this scope, Pieper and Korpel (1983) suggested digitally merging several images of the same scene captured with different focus setting for generating an all-in-focus (AIF) image. In their research, Pieper and Korpel (1983) merged the in-focus regions of the differently focused images based on a pixel-wise criterion in order to yield the *extended depth-of-field*. Subsequently, Sugimoto and Ichioka (1985) suggested applying a sharpness criterion in a local support window. The generation of AIF images by digital composition will be referred to as *focus stacking*. Fig. 2.7 illustrates focus stacking for a focus sequence of 13 frames.

State-of-the-art algorithms proposed in the literature to compute the AIF image (focus stacking algorithms) can be broadly organized into four main families: methods based on the spatial frequency, image pyramids, defocus modeling and wavelet transforms. Methods based on the *spatial frequency* usually apply a sharpness measure or focus measure in order to identify the pixels with highest information content in each frame (Li et al., 2001; Bilcu et al., 2009; Aslantas and Kurban, 2010; Tian and Chen, 2010). In turn, methods based on *image pyramids* usually perform a multi-scale decomposition of the image in order to identify the pixels or image regions with highest information content at different scales (Burt and Kolczynski, 1993; Zhang, 1999; Antunes et al., 2005). Alternatively, methods based on *defocus modeling* recover the AIF image under the assumption of a known point spread function model and then apply a filter designed to reverse its

effect (Subbarao and Choi, 1995; Kodama et al., 2006, 2007; Aguet et al., 2008). In order to work appropriately, the latter methods rely on an estimation of the parameters of the point spread function. Finally, methods based on the *wavelet transform* carry out a wavelet decomposition of the focus sequence. Image fusion is then performed in the wavelet domain by selecting the wavelet coefficients according to some criterion (Wang et al., 2003; Forster et al., 2004; Shirai and Ikehara, 2005; Tian and Chen, 2010; Baradarani et al., 2012). The wavelet transform can also be considered as an instance of a multi-scale decomposition, resembling the approaches based on image pyramids, although the coefficients are selected in the wavelet domain instead of in the spatial domain.

In general, most all-in-focus algorithms, with the exception of modeling-based methods, can be described through the following energy maximization scheme (Aguet et al., 2008):

1. An image stack, $I_k(x, y)$, is acquired by performing a focus sweep. The frames are captured so that they cover the whole focus range of interest.

2. Either a high-frequency measure or a focus measure operator is applied to each frame of the image stack. The operator is applied either in the space domain, frequency domain or wavelet domain depending on the all-in-focus algorithm used to compute the AIF image.

3. An index map is generated such that each position (x, y) keeps the index k of the frame with the largest frequency or focus measure for that position.

4. The AIF image is generated based on the index map. In the particular case of pyramid-based or wavelet-based methods, an inverse transformation is usually required.

An important drawback of the energy maximization scheme described above is that, in the presence of noise, the maximization of the focus measure will also increase noise in the final result. The effect of noise can be attenuated by applying a low-pass filter at the expense of image contrast. Alternatively, a noise-robust selective image fusion algorithm is presented in chapter 5 of this dissertation. The proposed algorithm is based on the analysis of the response of focus measure operators in non-ideal conditions and it allows the computation of low-noise all-in-focus images.

2.2.5 Shape-from-defocus

Since the work by von Helmholtz (1924), the role of accommodation and defocus in the perception of depth has extensively been assessed in the literature. More

recently, it has been experimentally shown that, in fact, the observed amount of defocus blur is an independent human pictorial depth cue by itself (Mather, 1996; Marshall et al., 1996). Similarly, in computer vision, defocus has been exploited in order to retrieve depth information by means of the *shape-from-defocus* (SFD) technique. In contrast to range-from-focus and shape-from-focus methods, which estimate depth by determining the location of the focus peak in a sequence of images, SFD aims at estimating depth directly from the amount of blur.

This principle can be intuitively explained as follows. Let $I_B(x, y)$ denote a blurred image generated from a source radiance, $I_S(x, y)$. From (2.1) and (2.5), this image can be computed as:

$$I_B(x, y) = I_S(x, y) * h_\rho(x, y), \tag{2.15}$$

where $h_\rho(x, y)$ is the blurring kernel.

Notice that from (2.4), the blurring radius ρ has a one to one correspondence with the point depth u_x. Therefore, if the acquisition parameters are known (lens focal length, lens-sensor distance and lens diameter), and a specific blur kernel model is assumed, the SFD problem can be interpreted as an inverse filtering or deconvolution problem. As a result, SFD can be applied to single images (Namboodiri and Chaudhuri, 2007). Notwithstanding, most practical approaches require two or more images for accuracy. The inverse filtering problem is an active research field in computer vision mostly exploited for image enhancement, denoising and deblurring. A detailed study of these techniques is beyond the scope of this dissertation. However, for the sake of completeness, previous research on this topic specifically tailored to or applied for solving the SFD problem is reviewed below.

Depending on the acquisition conditions and the available knowledge of the scene and the blur kernel, solving the inverse filtering problem is a non-trivial task. In the Fourier domain, the convolution in (2.15) can be used to compute the radiance of two defocused images, I_1 and I_2, as:

$$\mathcal{F}\{I_{B1}\} = \mathcal{F}\{I_S\}\mathcal{F}\{h_{\rho_1}\}, \text{ and } \mathcal{F}\{I_{B2}\} = \mathcal{F}\{I_S\}\mathcal{F}\{h_{\rho_2}\}, \tag{2.16}$$

where $\mathcal{F}\{\cdot\}$ is the Fourier transform operator, and ρ_1 and ρ_2 are the blur parameters corresponding to two different focus settings.

Under the assumption of a Gaussian model for the PSF, Pentland (1987) applied (2.16) in order to factor out the source radiance, yielding:

$$\mathcal{F}\{I_{B1}\} = \mathcal{F}\{h_{(\rho_2-\rho_1)}\}\mathcal{F}\{I_{B2}\} \tag{2.17}$$

For the case of images of blurred sharp edges, Pentland (1987) derived closeform expressions relating ρ_1 and ρ_2 in (2.17), thus allowing the estimation of

depth. Subsequently, Subbarao (1988) extended the application of (2.17) by using the power spectrum in order to obtain a more general solution, not limited to sharp edges. Spatial-domain solutions based on (2.17) were further developed by Hwang et al. (1989) and Subbarao and Surya (1994).

Notice that (2.17) prevents the need for estimating the image irradiance I_S. As an alternative, the direct deconvolution method requires the simultaneous estimation of both the image irradiance and the blur kernel. In general, the inverse filtering problem can be stated as (Favaro and Soatto, 2006):

$$\tilde{I}_S, \tilde{z} = \arg\min_{z, I_S} \left\| I(x,y) - \int\int h_z(x,y) I_S(x,y) \right\|^2, \tag{2.18}$$

where $I(x,y)$ is the observed image, $\tilde{I}_S$ and $\tilde{z}$ are the estimated scene radiance and depth-map, respectively. Note that the blur kernel has been expressed as a function of the scene geometry, z, instead of the blur parameter, ρ, by applying (2.4). The extension to deal with several images is straightforward. Nevertheless, (2.18) is a severely ill-posed inverse problem. In addition, as in typical deconvolution problems, a regularization term can be added to (2.18) in order to guarantee the smoothness of the obtained reconstruction (Vogel and Oman, 1998; Chan and Wong, 1998).

Early solutions to the inverse filtering problem in SFD can be traced back to Ens and Lawrence (1993). Under the assumption of a locally shift-invariant blur, Ens and Lawrence (1993) derived a matrix-based deconvolution framework for SFD. Later approaches under the same assumption have exploited different techniques such as moment filters (Xiong and Shafer, 1993; Watanabe and Nayar, 1998) or approximating the images in the spatial domain through simple discretizations (Subbarao and Surya, 1994; Rajagopalan and Chaudhuri, 1997; Favaro and Soatto, 2005). Some researchers have also relaxed the shift-invariant blur assumption and exploited diffusion (Favaro et al., 2008), linearized variational frameworks (Favaro, 2010) and Markov random fields (Rajagopalan and Chaudhuri, 1999).

As stated in chapter 1, defocus is not only a function of the camera's focus. Thus, some authors have tackled the problem of finding the optimum set of parameters for performing SFD (Rajagopalan and Chaudhuri, 1997). The application of SFD by changing the lens aperture and focal length has also been studied (Subbarao and Choi, 1995; Baba et al., 2001, 2006). The SFD problem has also been studied in the presence of blur due to camera shake (Favaro et al., 2004; Paramanand and Rajagopalan, 2012).

2.2.6 Focus calibration

The depth-of-field implies that, due to the limited resolution of real imaging systems, when the radius of the blurring circle, ρ, is below a negligible value, the target P in Fig. 2.5 can still be considered to be in focus. As a result, it is possible to tolerate a small focus error, as long as the resulting blur is below the maximum allowed blur, $\rho_{\max}$. More specifically, any target located between a near limit, u_n, and a far limit, u_f, around u is said to yield a negligible focus blur. Formally, the DOF corresponds to the distance range $u_f - u_n$. Using the thin lens model, it can be readily shown that (Pentland, 1987; Hwang et al., 1989; Xiong and Shafer, 1993):

$$u_n = \frac{uf^2}{f^2 + N\rho_{\max}(u - f)} \tag{2.19}$$

$$u_f = \frac{uf^2}{f^2 - N\rho_{\max}(u - f)}, \tag{2.20}$$

In order to obtain an effective control of focusing in applications such as autofocus, focus stacking, shape-from-focus and shape-from-defocus, it is necessary to know the limits of the DOF and its variations as a function of the current configuration of the camera. As previously stated, the near and far limits of the DOF can be found by applying (2.19) and (2.20), respectively. Notwithstanding, this requires an accurate knowledge of the parameters of the lens-camera system. On the one hand, the real physical values of the parameters of off-the-shelf conventional cameras are, at the best of cases, only known approximately. On the other hand, the maximum allowed blur, $\rho_{\max}$, is an empirical parameter defined according to the real pixel dimensions and the resolution of the system. As a result, some calibration methods have been devised in order to estimate the DOF limits and the blurring circle experimentally.

Starting from (2.4), Subbarao and Gurumoorthy (1988) expressed the blur radius ρ in the form:

$$\rho = mu^{-1} + c, \tag{2.21}$$

where m and c are camera-dependent constants.

In order to calibrate the constants in (2.21), the camera is set to a fixed configuration and a step target is placed at different positions from the camera $\{u_x | x = 1, 2, \cdots, X\}$, where X is total number of target positions used during the calibration process. For each target position, a blur radius, ρ_x, is found by fitting a *line spread function* to the blurred edge[4]. The parameters are then estimated by adjusting (2.21) to the obtained ρ_x vs. u_x curve. Subbarao exploited this procedure in order to perform depth retrieval through SFD. With this method, changing

[4]The line spread function is the analytical response of an optical system to a step edge

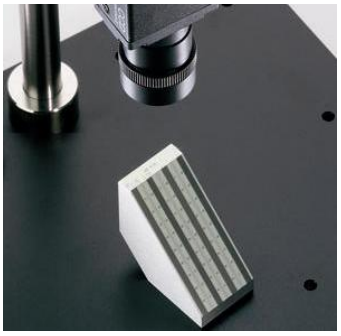


Figure 2.8: Gauging the depth-of-field. The DOF limits can be measured experimentally by means of a careful setting with a calibration gauge. Image courtesy of *Edmund Optics*. Used by permission.

any parameter of the camera (e.g., focus) requires repeating the calibration process. A similar method was proposed by Baba et al. (2001). In this latter case, the calibration procedure was extensively carried out for each camera setting by changing the focus, focal length and aperture of the camera. The advantage of the method proposed by Baba et al. (2002) is that the calibration parameters for each setting can be obtained with as few as two images.

A common industrial practice is to directly measure the DOF limits by means of *calibration gauges*. As opposed to the alternatives above, this approach requires as few as one single image for each camera setting. As illustrated in Fig. 2.8, the procedure consists in placing the calibration gauge in front of the camera, and measuring the distance range at which the blur is negligible. Again, the calibration procedure must be repeated if the focus (or any other parameter) of the camera is changed. Therefore, previous calibration procedures are suitable for cases when the camera is intended to work at a single (or very few) fixed settings. However, most conventional cameras (from cellphone cameras to professional photographic cameras) often require continuously adjusting focus when capturing images. As a result, those calibration procedures are rendered unpractical in these cases since they should be extensively repeated.

Alternatively, a practical and robust calibration procedure that allows describing the focus of the camera without explicit knowledge of its internal parameters is proposed in this dissertation. In order to develop this calibration method, the thin lens model is adapted for complying with two conditions: 1) The blur width should not depend on internal geometrical parameters (such as the sensor position v). Instead, it should depend on measurable external parameters such as the in-focus position, u. 2) The model implicitly accounts for the effects of the parameters of the system (focal length and f-number), without needing to know or estimate them explicitly, thus allowing an efficient calibration over the whole focusing range at once.

It is worth mentioning that, although the most common camera model is the thin lens model, alternative models with additional parameters also exist, such as the *thick lens* model (Horn, 1990) and the *pupil centric* model (Aggarwal and

Ahuja, 2002). These models include additional parameters for describing the projective geometry and the radiometric response of the system, but the basic operation of the focus mechanism remains the same.

2.2.7 Related trends

As illustrated in the previous sections, significant research efforts have been devoted to focus-related issues in the field of computer vision so far. Alternatively, there exist different approaches that exploit focus with the aid of special equipment tailored to specific applications. Although this dissertation limits the study of the focus phenomenon to conventional cameras, for future reference, those alternative approaches are briefly reviewed in this section.

From the very beginnings of the shape-from-defocus framework, Engelhardt and Knop (1988) exploited active illumination in order to perform real-time SFD. Subsequently, active depth recovery by projection of light patterns has been exploited by Watanabe and Nayar (1998); Nayar et al. (1996) and, more recently, by Lenz et al. (2012);

In recent years, new trends have evolved into the development and design of novel non-conventional cameras that exploit the focus cue in different manners. From the preliminary results by Adelson and Wang (1992) to the development of the first prototypes by Ng et al. (2005) of the so-called *plenoptic camera*, a novel field, computational photography, has increasingly gained the attention of researchers. Plenoptic cameras are modified devices that capture the light-field behind the lens, instead of the 2D projection that conventional cameras can capture (Bishop and Favaro, 2012). This enables new imaging functions such as digital refocusing (Ng, 2006) and 3D microscopy (Levoy et al., 2006) from single snapshots.

Within computational photography, cameras with flexible depth-of-field, which translate the sensing device during the image capture, have also been proposed (Kuthirummal et al., 2011); as well as multiple color-filtered aperture (MCA) cameras for multi-focus capture (Malik et al., 2007; Kim et al., 2012). A concept that consists of introducing coded patterns into the optic imaging path, namely *coded aperture*, can be traced back to (Fenimore and Cannon, 1978). This concept has been recently exploited for single-snapshot extended depth-of-field and depth-map computation (Veeraraghavan et al., 2007; Levin et al., 2007).

2.3 Preliminary experiments

The previous sections have outlined important concepts and relevant literature for understanding both the focus mechanism of conventional cameras and its applications. The aim of this section is to provide a critical review of some of these

concepts from a practical perspective. Special emphasis is given to different problems found in the acquisition of focus sequences and strategies for correcting them. The concepts presented in this section have partially been published in (Pertuz et al., 2010).

2.3.1 Defocus model limitations

In section 2.1, when the focus defect is considerable and diffraction effects are neglected, the geometrical optics approximation of the PSF of a defocused system corresponds to the pillbox function in (2.5). Alternatively, some researchers have suggested using a 2D Gaussian in order to take into account the effects of polychromatic illumination, lens aberrations and other defects. At this point, it is worth asking if the Gaussian approximation is indeed a suitable model for a defocused system.

A thorough review of the literature reveals that the assumption of a Gaussian blur model has widely been accepted and exploited for different applications. One of its main advantages is its mathematical tractability derived from the properties of the Gaussian function. For instance, the Gaussian PSF has successfully been exploited in computer vision for depth retrieval through SFD (Pentland, 1987; Subbarao and Surya, 1994), for proposing new focus measure operators (Yousefi et al., 2011), for computing the all-in-focus image in focus stacking (Kodama et al., 2007; Aguet et al., 2008) and for image restoration and deblurring (Cao et al., 2010; qing Qin, 2010; Orieux et al., 2010; Lai, 2011; Chen and Li, 2013; Paramanand and Rajagopalan, 2012). The Gaussian PSF has also been exploited for assessing the effect of defocus blur in human depth perception (Mather and Smith, 2002). In contrast, some authors have opted for more general models when estimating the PSF of optical systems (Szeliski, 2011; Williams, 1999; Joshi et al., 2008; Fergus et al., 2006; Ji and Wang, 2012) or for accurately assessing human visual acuity (Thibos, 2009; Watson and Ahumada, 2008). Based on this literature review, it can be remarked that assuming a Gaussian PSF has yielded satisfactory results for assessing image blur, both qualitatively and quantitatively. Notwithstanding, this assumption no longer applies when accurately assessing non-defocus aberrations of optical systems.

Note that both the Gaussian and the pillbox PSF are based on the geometrical optics approximation that neglects the effects of diffraction. In order to illustrate the implications of this simplification, the following experiment has been conducted. The accurate diffracted monochromatic PSF of a defocused system, $h_\lambda(x, y)$, has been computed by following (FitzGerrell et al., 1997). Sub-index λ has been added in order to explicitly indicate the dependence on wavelength. Since we are interested in the general case of polychromatic illumination, the monochromatic PSF is integrated along the wavelengths of the visible spectrum:

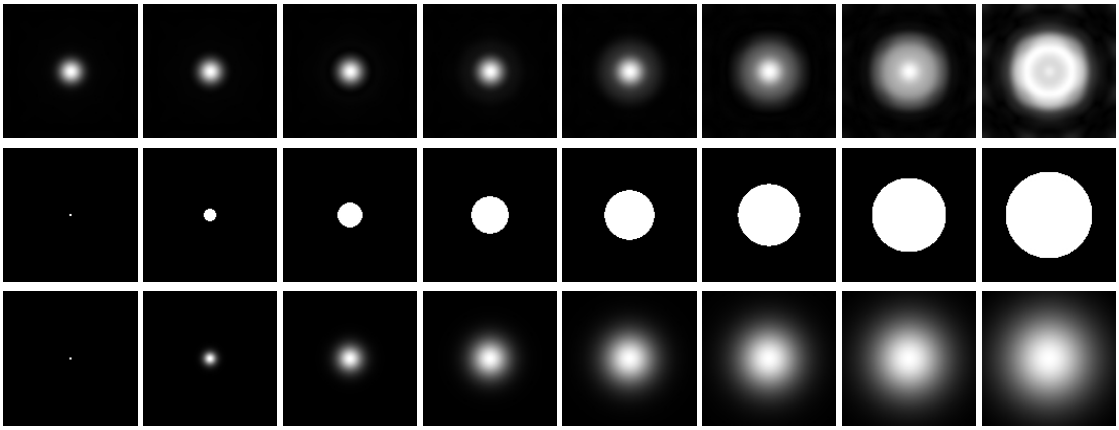


Figure 2.9: PSF of a defocused system as a function of the focus error, ϵ. Top row: accurate PSF taking into account the effects of diffraction (FitzGerrell et al., 1997). Middle row: geometric approximation in (2.5). Bottom row: Gaussian approximation in (2.6). Simulation parameters: $f = 35$ mm, $D = 15$ mm and $\lambda \in [380, 760]$ nm. Target at infinity.

$h(x, y) = \int h_\lambda(x, y)\, d\lambda$ by exploiting the superposition principle[5]. For comparison purposes, Fig. 2.9 plots the impulse response of a defocused system using the accurate PSF that takes into account the effects of diffraction (top row), the pillbox approximation (middle row) and the Gaussian approximation (bottom row), for different focus errors.

Fig. 2.9 suggests that, at least qualitatively, the geometric and Gaussian approximations describe reasonably well the effect of defocus for relatively large focus errors ($\epsilon > 11\mu$m). In contrast, the effects of diffraction impose a significant difference for small focus errors. This is in agreement with the results by Stokseth (1969), who stated that the geometrical approximation is similar to the exact diffraction PSF for large amounts of defocus. At this point, it is important to remark that, even if the Gaussian and pillbox functions are granted as acceptable models for badly defocused systems, there exist a conceptual "gap" near the focus position that can only be explained by diffraction.

A simple, yet complete, defocus model is desirable not only from a theoretical perspective but also in practice. In future chapters, this problem is tackled and we show that it is possible to predict the response of a defocused optical system with the geometrical optic approximation even at low defocus levels. In order to achieve this, it is necessary to consider the effects of both the optic of the system and the

[5]Strictly, since the quantum efficiency of the camera's sensor is a function of wavelength, the overall response of the camera to the illumination is also wavelength-dependent. As a result, this integral is weighted by the spectral response of the camera $C_r(\lambda)$. For simplicity, $C_r(\lambda) = 1$ has been used.

(a) (b)

Figure 2.10: Effect of field curvature. (a) An image of a fronto planar scene. The amount of blur on the center and the image borders is different due to the curvature field. (b) Depth-map obtained using SFF (raw depth-map without post-processing). z-axis: pixel depth (mm).

sampling process, which take place during the image formation process presented in the previous section. The benefit of this approach is that it takes advantage of the simplicity of the thin lens model while providing a consistent defocus model over the whole focusing range (chapter 4).

2.3.2 Field curvature

There are two important optical artifacts inherent to the acquisition process during the focus sweep, which can alter the results of focus-related applications: the field curvature and the pixel shift. As stated in section 2.1.1, the field curvature is one of the five monochromatic Seidel aberrations studied in geometrical optics. Ideally, an optical system projects the image of the observed scene on a plane, where it is recorded by a flat sensor (e.g., a CCD). However, in the presence of this aberration, the projected image departs from a flat surface leading to a spatial mismatch between the real image and the sensor. As a result, the sensor samples part of the space in front of or behind the sharp image, and the recorded image will thus be locally blurred. For illustration, Fig. 2.10a shows an image of a planar object captured with a camera with field curvature aberration. The difference in sharpness along the image field is evident. As shown in Fig. 2.10b, this aberration can lead to an erroneous depth estimation in applications such as SFF.

In order to compensate for the field curvature, the shape of the curvature itself must be estimated. For instance, Rosete-Aguilar (2000) exploited the see-saw diagram introduced by Burch (1942) in order to compensate for the field curvature in basic telescope configurations such as the Cassegrain, the Dall-Kirkham, the Pressman-Camichel and the Ritchey-Chretien telescopes. In astronomy, the blurring of distant point light sources, such as stars, can be exploited in order to perform empirical estimations of the field curvature by means of quality met-

rics such as the RMS spot size, the encircled energy and the edge transition width, among others (Vaughnn and Mark, 2006). Intuitively, these methods work by measuring the spread of the images corresponding to point light sources (the stars) and relating this measure with the amount of curvature along the image field. There are already modern software packages, such as *CCD Inspector*, which can automatically perform empirical estimations of the field curvature in telescopes based on these image metrics. Unfortunately, these approaches are not tailored to conventional cameras. As a result, focus-based shape estimation approaches use an *ad hoc* calibration procedure that consists in fitting a surface Z_{fit} to the depth-map obtained from a fronto parallel calibration plane (Nair and Stewart, 1992). Strictly, the field curvature can be associated with the Petzval curvature, the tangential, sagittal or medial curvature, in which case the shape is either spherical or paraboloidal (Vaughnn and Mark, 2006). Since the 3rd order approximation of a sphere is also paraboloidal, Z_{fit} can be modeled as a 2D paraboloid. The effect of the field curvature is then compensated for in a post-processing step by subtracting Z_{fit} to the depth-maps obtained without compensation.

Remarkably, the lenses of most conventional cameras are provided with compensators that reduce the effects of the field curvature. This is often the case for compact digital cameras, SLR cameras and surveillance cameras. This effect is only appreciable in devices with simple optics such as some webcams and cell phone cameras. For instance, Fig. 2.10 corresponds to a web camera. In the sequel, it is assumed that this effect is either negligible or has been corrected by the described procedure.

2.3.3 Pixel shift

As stated in section 2.1.3, when a camera changes its internal configuration (for instance, by performing a focus sweep), the intrinsic parameters of the camera may vary mostly due to marginal changes in the magnification. As illustrated in Fig. 2.11, in addition to a change in the focus level, the change in the focus setting results in a local shift of image features. The side effect of focus (accomodation) on the image magnification was first reported by von Helmholtz (1924) and studied by Biersdorf and Baird (1966) in human vision. This *feature shift* can represent a problem in applications such as focus stacking and shape-from-focus, whereas, in the generation of the focus measure vector, it is assumed that the image coordinates corresponding to each scene point remain constant during the acquisition process.

The side effect of focusing in magnification has been noticed by many researchers and some attempts have been made to address the feature shift problem. Thus, Nair and Stewart (1992) proposed the use of larger support windows for the computation of the focus measure operator at the cost of spatial resolution. Darrell and Wohn (1988) proposed to construct a distortion map by taking images

(a) (b)

Figure 2.11: Effect of magnification change on feature shift (a) Focus at $u = 4.6$ m. (b) Focus at $u = 8.7$ m. Images captured with a Nikon D90 DSLR camera.

of a test pattern and tracking some key points. In that work, the information of the distortion map is used to predict pixel shift due to changes in focus. A similar approach was used by Hasinoff et al. (2009) but, instead of an explicit distortion map, a quadratic distortion term as a function of focus was incorporated to the geometric model of the camera. The drawback of these two methods is that they depend on the accuracy to track the key points of the pattern and also require an extensive calibration procedure. Alternatively, Watanabe and Nayar (1995) proposed the use of a telecentric lens system in which magnification is kept constant relative to focus variations. Willson (1994) proposed to compensate for magnification changes due to focusing by means of zoom. In this approach, a zoom value must be determined to compensate for the changes in image magnification for every focus position. Both in (Watanabe and Nayar, 1995) and (Willson, 1994), either complex controllable optics or extensive lens calibration procedures are required.

CHAPTER 3

Focus measure

A critical step in many focus-related tasks is the estimation of the relative focus level or *focus measure*. In computer vision, this is achieved by means of the so-called focus measure operators or focus measure algorithms. For practical reasons, it is of up-most importance to understand the working principles of focus measure algorithms as well as the imaging factors that have an impact on their performance. In this chapter it is shown that, indeed, focus measure operators respond differently to various imaging factors, such as noise level and support window, according to their working principle. The contributions presented in this chapter are applicable to the development of new focus measure operators or for guiding their application.

The performance of up-to-date focus measure operators is analyzed in this chapter. Section 3.1 introduces the problem of focus measure performance and reviews relevant previous work. Section 3.2 provides a summary of different focus measure operators. A methodology for analyzing the performance of the studied focus measure operators is presented in section 3.4. The obtained results and the concluding remarks are presented in sections 3.5 and 3.6, respectively.

3.1 Introduction

Recent perceptual experiments have shown that the human visual system is capable of obtaining surprisingly high-precision estimates of defocus even at heterogeneous natural viewing conditions (Burge and Geisler, 2011). In computer vision, the development of algorithms based on image processing for focus measurement is challenging due to the number of factors that can influence the outcome of a focus measure operator. According to chapter 1, the degree of blurring and, therefore, the amount of focus of an imaged scene depends on different controls of the acquisition process, such as the lens aperture, focal length, scene geometry and focus setting of the camera. Notwithstanding, due to the characteristics of digital images and the constraints associated with their discrete nature, an additional set of *image-dependent* factors also have an impact on the perception of blur in computer vision. In particular, the image-dependent factors considered in this dissertation include image noise, contrast, saturation and the support window of the focus measure operator. As a result, a practical assessment of the performance of focus measure operators and how they respond to the aforementioned factors is of great interest.

Since the effects of the focus controls can be addressed from a theoretical perspective (chapter 4), this chapter presents a thorough analysis of the performance of focus measure operators as a function of the image-dependent factors. In order to be able to analyze the results from a practical viewpoint, the experiments are performed for a specific application field: shape-from-focus (SFF). In particular, SFF has been selected among the different focus-related applications presented in chapter 2, since it allows assessing the performance of focus measure operators directly by measuring the error of the estimated pixel depth. Unlike autofocus, where the focus measure operators are applied over relatively large regions, SFF involves small support windows for the computation of pixel-wise focus measures, which represents a more challenging scenario. In addition, in contrast to shape-from-defocus, the performance of SFF is less sensitive to the particular image frames and the inverse filtering model used to perform depth estimation, and does not require the calibration of camera-dependent parameters. As a result, the final reconstruction error can directly be attributed to inaccuracies in the estimation of the focus measures.

Previous work

Most comparative studies about focus measure operators have been carried out for autofocusing (AF) in microscopy (Firestone et al., 1991; Santos et al., 1997; Sun et al., 2004; Russell and Douglas, 2007). In particular, Groen et al. (1985)

proposed to apply the focus measure operator to an image stack and to compute the corresponding focus function. Some ideal properties of the focus function are then analyzed in order to determine its suitability for AF. Specifically, a focus function should be a smooth unimodal curve (with a single maximum), with a sharp peak at the exact location of best focus. Following Groen's work, Firestone et al. (1991) ranked different focus measure operators according to four features of the focus function: accuracy (deviation of the focus position from its correct value), range (height of the focus function), number of false maxima, and the function's width (measured at 50% of its peak value). Santos et al. (1997) complemented this methodology by also considering the execution time among the key features taken into account for ranking the considered operators. Subsequently, Sun et al. (2004) and Xie et al. (2007) performed an evaluation of different focus measure operators using a similar approach, but additionally including the *noise level* (energy of the local false maxima) and resolution (global distribution of the focus function) among the features of the focus function.

Unfortunately, to the best of our knowledge, no experimental results have been provided supporting the claim that, excluding accuracy, the features of the focus function (range, width, noise level, etc) are indeed a predictor of the performance of a focus measure operator in focus-related tasks, such as depth estimation or focus stacking. In addition, as previously stated, the focus measure is used to determine the position of the best focused image by applying the operator to a large region of interest in AF. In contrast, in applications such as shape-from-focus, shape-from-defocus and focus stacking, the focus measure must be estimated for every pixel, with the focus measure operator being applied using a small support window. Therefore, the results of comparative studies about focus measure operators applied to AF can be generalized in a very limited way. A more general theoretical method for assessing the uncertainty of various focus measure operators as a function of gray-level noise was proposed by Subbarao and Tian (1998).

For shape-from-focus in microscopy, Malik and Choi (2007) ranked the performance of five depth-map estimation techniques under different illumination conditions and window sizes. Although that particular work provides interesting experimental results (section 3.6), the discussion focuses on the reconstruction stage of shape-from-focus, instead of the focus measure operators. Moreover, in microscopy imaging, there are two important differences with respect to conventional cameras: firstly, focusing in optical microscopes is achieved by changing the relative position between the object and the sensing device (by moving the stage of the microscope), while keeping the optics fixed. Alternatively, in conventional photography, focusing is achieved by changing the internal configuration of the lens, with subsequent effects on the optics properties of the system. Secondly, the imaging conditions in microscopy are different due to the high magnifications,

very short working distances and shallow depth-of-field, as well as the controlled illumination. In addition, beyond an absolute ranking of focus measure operators according to their performance, we are rather interested in how they respond to different factors according to their working principles. The findings of previous researchers and their relationship with findings in this chapter will be discussed in more detail in section 3.6

3.2 Focus measure operators

A wide variety of algorithms and operators have been proposed in the literature to measure the degree of focus of either a whole image or an image pixel depending on the application. In order to facilitate an exposition of the working principles of the focus measure operators studied in this chapter, they have been grouped into six broad families: gradient-based, Laplacian-based, wavelet-based, statistics-based, DCT-based and miscellaneous operators. This section presents a brief description of each family. Notice that some of the operators studied were originally devised for autofocusing applications tailored to measuring the focus level of a whole image region. Therefore, they had to be extended and adapted in order to allow a pixel-wise focus measure. In addition, in order to keep a global perspective on the concepts behind the operators and facilitate the comprehension, the operators are not presented individually but within the scope of the corresponding family. A detailed exhaustive description of each focus operator, as well as their parameters and implementation details can be found in appendix A. Table 3.1 summarizes the abbreviations used in this dissertation to refer to the different focus measure operators. Operators with abbreviations in bold-face have been adapted in this work to perform pixel-wise focus measure estimation.

Gradient-based operators. This family groups the focus measure operators based on the gradient or approximations of the first derivatives of the image. These algorithms follow the assumption that focused images present sharper edges than blurred ones. Thus, the energy of the gradient can be exploited in order to estimate the degree of focus. This principle is exploited by the widely known Canny edge detector (Canny, 1986). The operators based on this principle are expected to work properly as long as the imaged scene is highly-textured. However, it is important to remark that this is a common restriction for most focus measure operators.

In the frequency domain, the gradient operator can be interpreted as a high-pass filtering of the image. On the one hand, this provides a sensitive response to defocus, which in turn corresponds to a low-pass filtering. On the other hand, a well-known issue is the noise sensitivity of gradient-based schemes, specially at small scales (Bergholm, 1987). The operators corresponding to this family are abbreviated as GRA* in table 3.1.

Table 3.1: List of focus measure operators and their abbreviations

Focus operator	Abbr.	Focus operator	Abbr.
Gradient energy	GRA1	Gray-level variance	STA3
Gaussian derivative	**GRA2**	Gray-level local variance	**STA4**
Thresholded absolute gradient	GRA3	Normalized gray-level variance	STA5
Squared gradient	GRA4	Modified gray-level variance	**STA6**
3D gradient	GRA5	Histogram entropy	STA7
Tenengrad	GRA6	Histogram range	STA8
Tenengrad variance	**GRA7**	DCT energy ratio	**DCT1**
Energy of Laplacian	LAP1	DCT reduced energy ratio	**DCT2**
Modified Laplacian	LAP2	Modified DCT	**DCT3**
Diagonal Laplacian	LAP3	Absolute central moment	**MIS1**
Variance of Laplacian	**LAP4**	Brenner's measure	**MIS2**
Laplacian in 3D window	LAP5	Image contrast	**MIS3**
Sum of wavelet coefficients	**WAV1**	Image curvature	**MIS4**
Variance of wavelet coefficients	**WAV2**	Hemli and Scherer's mean	**MIS5**
Ratio of the wavelet coefficients	WAV3	Local Binary Patterns-based	**MIS6**
Ratio of curvelet coefficients	WAV4	Steerable filters-based	**MIS7**
Chebyshev moments-based	**STA1**	Spatial frequency measure	**MIS8**
Eigenvalues-based	**STA2**	Vollath's autocorrelation	**MIS9**

Laplacian-based operators. Similarly to the previous family, the goal of these operators is to measure the amount of edges present in the images, although through the second derivative or Laplacian. The image Laplacian is also a widely-known basic image processing tool used for edge detection and image enhancement (Torre and Poggio, 1986; Gonzalez and Woods, 2008). Its downside is its increased sensitivity to noise as compared to the image gradient (Haralick, 1984). The operators corresponding to this family are abbreviated as LAP* in table 3.1.

Wavelet-based operators. The wavelet decomposition of an image can be interpreted as a simultaneous frequency and scale-space analysis (Mallat, 1989). Fig. 3.1 illustrates the computation of the 1st-level discrete wavelet transform (DWT) coefficients by means of the *two-channel filter bank* scheme (Strang and Nguyen, 1996). In this scheme, the image is decomposed into 4 sub-images by means of a high-pass filter, G_H, and a low-pass filter, G_L, which operates on the source image row-wise and column-wise alternately. This yields three detail sub-bands that emphasize the horizontal variations (W_{LH}), the vertical variations (W_{HL}) and the diagonal variations (W_{HH}), and a coarse approximation image (W_{LL}). In order to keep the number of total pixels constant, the computation of the wavelet coefficients implies downsampling in order to halve the size of the coefficient sub-bands (not shown in Fig. 3.1). This process can be further repeated on the coarse approximation image in order to add more levels to the decomposition.

Notice that, by following a similar reasoning as in previous families, the energy of the detail sub-bands can be used for estimating the degree of focus of an image, since they are related to the highest frequencies of the image (Gopinath et al., 1994). From a spatial-domain perspective, the wavelet transform can be interpreted as a multi-resolution representation of an image. This fact makes it

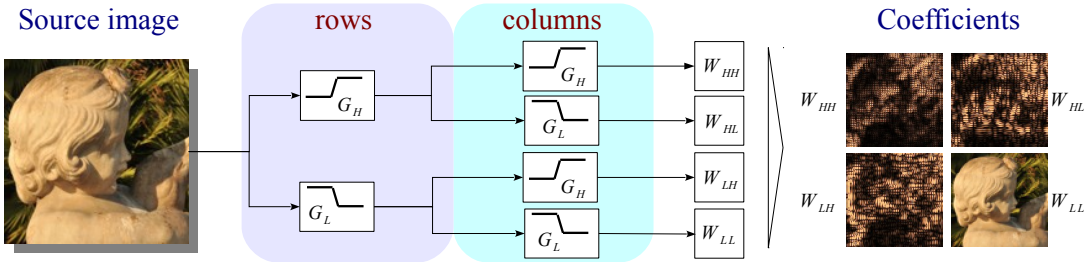


Figure 3.1: Wavelet decomposition for the computation of the discrete wavelet coefficients (DWT). The colormaps of the sub-bands have been modified for display purposes.

suitable for addressing the support window issue in the application of the focus measure operators (that is, the problem of selecting an appropriate support window size). This fact has been exploited not only for focus measurement but also for focus stacking (Forster et al., 2004; Wang et al., 2003) and image compression for the JPEG2000 standard (Taubman and Marcellin, 2002). The operators corresponding to this family are abbreviated as WAV* in table 3.1.

Statistics-based operators. In the spatial domain, the effect of defocus can be assessed from its effects on the textures of the imaged scene. In turn, statistical operators have proven to be quite successful as texture descriptors (Petrou and Sevilla, 2006). Intuitively, a defocused image can be interpreted as a texture whose smoothness increases for increasing levels of defocus.

In real imaging conditions, in the presence of different noise sources, statistical moments such as the variance and Chebishev moments, the energy of the principal components, etc, are robust texture descriptors. In fact, interpreting the image as a noisy 2D statistical process has been exploited in image restoration through inverse filtering, Wiener restoration and image denoising, among others (Berriel et al., 1983; Pratt, 2007). The operators corresponding to this family are abbreviated as STA* in table 3.1.

DCT-based operators. The discrete cosine transform (DCT) can be interpreted as an alternative to the Fourier transform for the representation of signals in the frequency domain. One of its main characteristics is its ability to pack most of the information of the input signal in the lowest coefficients of the transform. This characteristic is referred to as the energy compaction property. Thus, Reininger and Gibson (1983) empirically showed that the distribution of the DCT coefficients follows the Laplace distribution. Subsequently, this fact has been demonstrated theoretically by Lam and Goodman (2000). The compaction property can be exploited for achieving lossy compression, for instance, of image and video signals (Wallace, 1992; Sikora, 1997). In the space-domain, the DCT coefficients can be interpreted as an estimator of the image sharpness. For instance, as noted by Baina and Dublet (1995), the sum of the AC components of the DCT is equal to

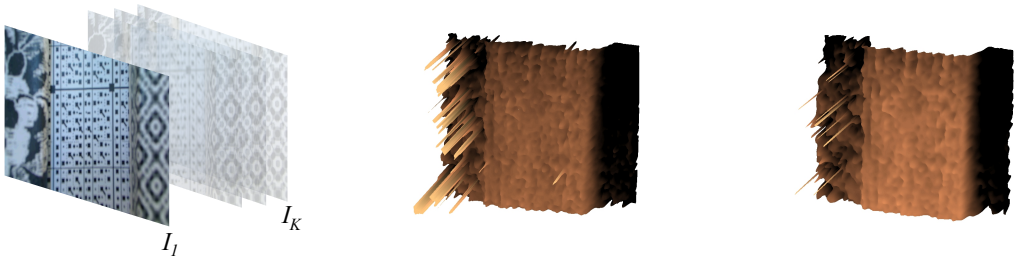


Figure 3.2: Compensating for image magnification shift. (a) First and last frames of focus sweep. (b) Depth-map obtained using SFF without shift compensation. (c) Depth-map obtained using SFF with shift compensation (z-axis: pixel depth in milimeters).

the variance of the image intensity and can, therefore, be used as a focus measure. Being part of many popular video and image formats, the main motivation for using the DCT for focus measure has been the reduced cost of its computation. Notwithstanding, in these formats, the DCT coefficients are typically computed in fixed support windows of 8×8 pixels. As a result, DCT-based operators have mostly been applied to autofocusing. The operators corresponding to this family are abbreviated as DCT* in table 3.1.

Miscellaneous operators. This family groups operators that do not belong to any of the previous five groups. The operators in this group are based on different concepts, such as the image contrast, local binary patterns and steerable filters, among others. The operators in this family are abbreviated as MISC* in table 3.1.

3.3 Magnification shift compensation

As previously stated in chapter 2, a side effect of focusing is a change in the magnification of the system. Alternatively to the extensive calibration procedures or especialized optics presented in section 2.3.3, this effect can be compensated for by applying the focus measure operator in neighborhoods that adaptively change their location as a function of image shift. For this purpose, many registration techniques including point matching, image correlation and FFT can be exploited. In this thesis, we use phase correlation due to its simplicity and robustness to noise, and since it can be adapted in order to tolerate blur (Ojansivu and Heikkila, 2007). A more detailed description of shift estimation using phase correlation can be found in Foroosh et al. (2002).

The phase correlation method is based on the translation property of the Fourier transform, which can be summarized as follows (Foroosh et al., 2002; Raj

and Staunton, 2007): Let $I_1(x, y)$ and $I_2(x, y)$ be two images that differ from a displacement (x_0, y_0). According to the Fourier translation property, their corresponding Fourier transforms F_1 and F_2 will be related by:

$$F_2(u, v) = F_1(u, v) \cdot e^{-j2\pi(ux_o + vy_o)} \tag{3.1}$$

Therefore, if the phase difference component of their Fourier transforms is isolated, then its inverse will correspond to a Dirac delta function centered at (x_0, y_0). The phase difference in (3.1) can be obtained by computing the normalized cross power spectrum of F_1 and F_2.

Taking into account the image feature shifts due to magnification effects, the focus measure operator is applied in two steps:

1. Let I_k and I_{k+1} be two consecutive images in a given image stack. The horizontal and vertical shifts $(\delta x_k, \delta y_k)$ for every pixel (i_k, j_k) between I_k and I_{k+1} are computed using phase correlation in an $M \times N$ window centered at that pixel. A Hamming window is used in this step to reduce the effects of sub-image edges.

2. The shift-compensated focus measure vector corresponding to every pixel is defined as, $\varphi_{i,j} = (F_1(i_1, j_1), ...F_K(i_K, j_K))$, such that:

$$\begin{cases} (i_1, j_1) = (i, j) \\ (i_{k+1}, j_{k+1}) = (i_k + \delta x_k, j_k + \delta y_k) \end{cases}$$

From this point, focus stacking or shape-from-focus frameworks can proceed normally by operating on the shift-compensated focus measure vectors.

In order to illustrate the effect of the feature shift compensation, Fig. 3.2 compares the depth-maps obtained by means of SFF with and without shift compensation. It can be seen that, mainly at the periphery of the image field, the magnification shift has a negative impact on the estimated depth. Again, the change in magnification as a function of focus depends on the particular imaging device. The shift compensation approach described above yields improvements of up to 21% in the root mean squared error (RMSE) of the obtained depth-map. Preliminary tests were also carried out with different well known optical-flow algorithms, such as the Black-Anandan, Lucas-Kanade and Horn-Schunck methods, and the best results were obtained with phase correlation, arguably due to its robustness to the effect of blur.

3.4 Comparative methodology

The focus measure operators introduced in the previous section have been applied to sequences of both synthetic and real images in order to obtain the depth-maps

of different scenes through shape-from-focus. The implemented depth estimation routine has been the classic SFF framework in (Nayar and Nakagawa, 1994), without the post-processing stage (the output depth-map is not regularized or filtered). In order to assess the robustness of the evaluated focus measure operators, the procedure has been repeated under different factors: image noise level, size of support window, image contrast and saturation. The test procedure can be summarized in three steps:

1. For each image sequence of the test set, generate a depth-map, $z(x, y)$, by means of SFF using each of the focus measure operators.

2. Compare each obtained depth-map with the ground-truth, $z_G(x, y)$, corresponding to each sequence of the test set. The obtained results are compared and analyzed according to specific *quality measures* (section 3.4.2).

3. Repeat the previous steps by adding noise to the test sequences and modifying the contrast, saturation and support window size.

In this section, the image set used to test the evaluated focus measure operators is first described. Afterwards, the evaluation procedure utilized to compare the performance of each operator is presented. Finally, the methodology to assess the robustness of each family to noise, contrast, saturation and support window size is described.

3.4.1 Image sequences

In order to provide test data for different imaging conditions, a set of twelve image sequences from three different imaging devices has been used: a webcam, a surveillance camera and a simulated camera. The characteristics and properties of each group of sequences are briefly described below. For illustration purposes, Fig. 3.3 shows the all-in-focus images corresponding to sequences captured with the different acquisition devices. As will be shown in the next sections, the smoothness and quality of the obtained depth-maps change depending on the acquisition conditions. Therefore, a diversified test set is important in order to avoid biasing the results due to the particular characteristics of the selected imaging device[1].

Surveillance camera. A group of 4 image sequences with 50 frames of 640×480 pixels captured with a Sony SNC-RZ50P surveillance camera has been considered. The focus sweeps were performed at maximum focal length ($f = 91$ mm) to obtain minimum depth-of-field. For all the scenes, the imaged objects are

[1]Some of the focus sequences are publicly available online at www.sayonics.com/downloads.html.

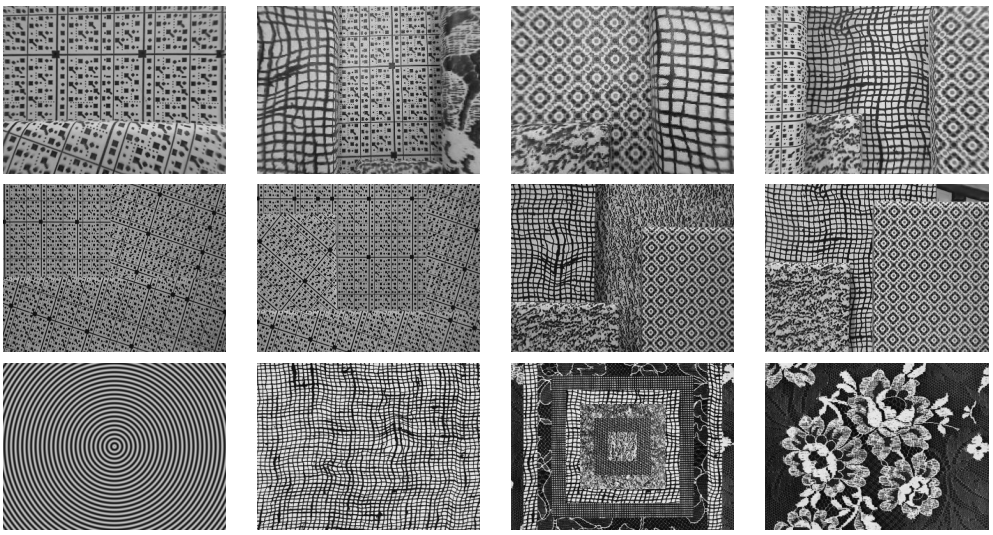


Figure 3.3: All-in-focus images for each focus sequence of the test set. Sequences from a webcam (top row), sequences from a surveillance camera (middle row) and simulated sequences (bottom row).

within a range of 36.8 cm and the focus sweep is performed by moving the in-focus position in a range of 188 cm around that distance. These sequences are characterized by low optical artifacts (vignetting, distortion and spherical aberration), as well as for a small image shift.

Webcam. A group of 4 image sequences with 51 frames of 640×480 pixels captured with a Logitech Orbit AF camera has been included. The considered focus sweep was between 11.90 mm and 81.0 mm. These sequences present some optical artifacts, in particular, radial distortion and vignetting, as well as a greater image shift and spherical aberration. The image shift and the spherical aberration have been compensated for by means of the methods described in section 3.3.

Simulated camera. A group of 3 sequences with 25 frames of 625×625 pixels and a sequence of 60 frames of 360×360 pixels synthetically generated have been considered. Defocus was simulated for a 3.3 mm focal length camera, focusing between 50 mm and 200 mm. For all the synthetical scenes, the imaged surface is between 100 mm and 150 mm away from the simulated camera. This camera represents an ideal imaging case without optical artifacts (only the defocus effect is considered). The details of the algorithm used to simulate defocus and generate synthetical focus sequences can be found in Appendix B.

3.4.2 Evaluation procedure

In order to compare the performance of different focus measure operators, a measure to evaluate the quality of the obtained depth-maps must be utilized. For this purpose, an objective error measure that compares the ground-truth, z_G, with the obtained depth-map, z, is defined from the root mean square error as:

$$E_{rms} = \sqrt{\frac{1}{MN}\sum_{(x,y)} |z_G(x,y) - z(x,y)|^2}, \tag{3.2}$$

where $M \times N$ is the size of the image in pixels.

Note that the root mean square error is an objective quality measure suitable for experiments performed under similar imaging conditions (i.e., from the same acquisition device, image noise level, contrast, saturation and support window), using different focus measure operators. In particular, since the value of E_{rms} is not normalized and depends on the units of the compared variables, the error measures obtained for a given image sequence cannot be directly compared to those of another sequence. In order to avoid biasing the overall results towards a particular experiment and emphasize on the *relative performance*, a relative quality measure, Q_r, is defined for each experiment as:

$$Q_r = \frac{E_{\max} - E_{rms}}{E_{\max} - E_{\min}}, \tag{3.3}$$

where $E_{\max}$ and $E_{\min}$ are the maximum and minimum error measures obtained for any of the focus measure operators in that experiment, respectively.

The relative quality in (3.3) can be interpreted as a standardized performance measure that assigns a value of 1 to the operator that yields the best performance, and 0 to the operator with the highest error measure in a particular experiment. Q_r has the advantage that its value only depends on the relative performance of the compared operators, thus being independent of the scales and units of both the ground-truth and the obtained depth-maps. This behavior is suitable for finding *trends* in the behavior of the focus measure operators as a function of the imaging factors, whereas it is not suitable for finding absolute rankings. For illustration purposes, tables 3.2a and 3.2b show the evaluation procedure for two different image sequences. For the sake of clarity, only three operators have been included in these tables[2].

Table 3.2a corresponds to an image sequence obtained with the webcam, whereas table 3.2b corresponds to a sequence from the surveillance camera. It can be clearly appreciated that the error values obtained by the operators differ in at least one

[2]No special preference is given to any of these operators. They were selected for this example due to their similarity in both definition and performance.

Table 3.2: Relative performance. (a) Experiment with sequence from the webcam. (b) Experiment with sequence from the surveillance camera.

(a)

Operator	E_{rms} (mm/pixel)	Q_r
WAV1	1.76	1.00
WAV2	2.12	0.10
WAV3	2.16	0.00

(b)

Operator	E_{rms} (mm/pixel)	Q_r
WAV1	51.1	1.00
WAV2	54.8	0.36
WAV3	56.9	0.00

order of magnitude between both sequences. For instance, the WAV1 operator yielded an E_{rms} of approximately 1.8 mm/pixel for the first sequence, and 51 mm/pixel for the second. This is not unexpected, since the first sequence corresponds to a focus sweep of 69 mm, whereas the second sequence corresponds to a sweep of 1880 mm. Therefore, such a difference in the scales and conditions of the experiments could lead to an erroneous interpretation of the results. In contrast, the values in the third column of both tables Q_r are dimensionless, although they clearly differentiate the relative performance of the different focus operators.

3.4.3 Imaging factors

The performance of the evaluated focus measure operators has been assessed by taking into account the effect of four different imaging factors: image contrast, image saturation, image noise and the size of the support window, as described below.

Window size. In SFF, a focus measure operator is applied to each image pixel by processing a small neighborhood, or support window, around it. The nature and amount of image information and, hence, the size of the support window can strongly affect the performance of a focus measure operator. Malik and Choi (2007) addressed the problem of determining the optimum window size for the application of focus measures for shape recovery. They observed that increasing the size of the support window can lead to an erroneous estimation of depth due to over-smoothing effects. On the other hand, as noted by Marshall et al. (1996), small windows increase the sensitivity to noise and the problem of image occlusion blur at sharp depth discontinuities. In order to evaluate the effect of the window size on the performance of the focus measure operators, nine different sizes between 3×3 and 17×17 pixels have been considered .

Image noise. Recall from chapter 2 (section 2.1.2) that a real digital image can be modeled as an ideal image, I, plus different noise components as:

$$I_n = g(I + n_s + n_c) + n_q, \tag{3.4}$$

where I_n is the noisy image, $g(\cdot)$ is the camera response function (CRF), n_s is the irradiance-dependent noise component, n_c is the independent noise, and n_q, is the quantization and amplification noise. In order to assess the effect of noise on the performance of the operators, the experiments have been repeated by increasingly adding noise to the original focus sequences according to (3.4). Following (Liu et al., 2008), n_q has been neglected and, n_s and n_c are modeled as Gaussian noise with zero mean and variances $Var(n_s) = I \cdot \sigma_s^2$ and $Var(n_c) = \sigma_c^2$, respectively. The focus measure operators have been evaluated with ten different noise levels, assuming an identity function for the ideal CRF.

Image contrast. The contrast of the captured image is another feature related to the image content that can affect the performance of a focus measure operator. Low contrast images usually contain smooth edges, thus increasing the difficulty to determine the relative degree of focus. Moreover, an operator too sensitive to variations in the image contrast will exhibit a variable behavior over the image field in the presence of some image aberrations such as vignetting. In order to assess the robustness of the different operators to reductions of image contrast, the experiments have been repeated by pre-processing the image sequences in order to reduce their contrast. In particular, for every image sequence, contrast was reduced by compressing their histograms through the following histogram equalization transfer function:

$$I_c(x, y) = c(I(x, y) - 128) + 128, \tag{3.5}$$

where $I_c(x, y)$ is the new image intensity of pixel $I(x, y)$ and c is the histogram compression ratio. This equation allows for a linear compression of the image histogram around its center for gray-levels between 0 and 255. In (3.5), the slope c of the transfer function is reduced in order to decrease the contrast of the image. This operation must be performed in unsigned integer format to achieve a real compression of the histogram instead of a simple scaling of the gray-level values.

Image saturation. This factor can also affect the performance of focus measure operators. In this work, image saturation has been evaluated by adding a constant offset to the original image:

$$I_s(x, y) = I(x, y) + S, \tag{3.6}$$

where $I_s(x, y)$ is the saturated pixel at coordinates (x, y) and S is the saturation level. The values of S are between 0 and 128 in order to obtain a saturation level from 0% to 50%. Again, it is assumed that image gray levels are coded in unsigned integer format and values above 255 are set to 255.

Table 3.3: Average computation time, t, of evaluated focus measure operators for all the considered image sequences.

Method	t (ms)	Method	t (ms)	Method	t (ms)
GRA3	5.60	MIS9	11.0	WAV2	88.00
GRA4	5.90	MIS8	12.0	GRA5	111.0
MIS9	6.00	GRA1	15.0	WAV3	125.0
DCT3	6.30	LAP4	15.0	LAP5	173.0
MIS2	7.10	MIS4	17.0	STA7	388.7
LAP1	7.10	STA8	17.1	MIS6	540.0
LAP2	7.20	STA3	18.0	STA1	6490
GRA2	7.30	STA4	18.0	STA2	6770
GRA6	9.70	GRA7	18.0	DCT2	8640
STA5	10.0	MIS7	22.0	DCT1	8830
LAP3	10.0	STA4	26.0	MIS1	10100
MIS5	10.7	WAV1	55.0	MIS3	12480

3.5 Experiments and discussion

All the evaluated focus measure operators were implemented in MATLAB and applied to the 12 focus sequences previously described in section 3.4.1. The experiments were conducted in two stages. In the first one, the overall performance of all focus measure operators when applied to all the focus sequences of the test set was evaluated. According to their performance, a sub-group of operators was pre-selected. At the second stage, the pre-selected group of operators was evaluated by changing the imaging factors, as described in the previous section.

3.5.1 Overall performance

For comparison purposes, all the operators were implemented and tested under the same platform. Table 3.3 summarizes the mean computational time obtained for every focus operator for 640×480 images on a Pentium IV quat core at 2.5 GHz. It is important to highlight that the performance of some operators highly depends on their particular application. For instance, the DCT-based operators were originally proposed to exploit the information inherent to some video and image formats. Therefore, the results of table 3.3 are provided for future reference, but should be interpreted accordingly.

As stated in section 3.1, in addition to accuracy (which in this case refers to the error E_{rms}), some researchers have utilized different features of the focus function in order to evaluate the performance of focus measure operators in autofocus. In order to determine if these features are indeed a robust predictor of the performance of focus measure operators for the computation of pixel-wise focus measure and, more specifically, for the computation of depth-maps via shape-from-focus, the following experiment was carried out: for each operator, a sample of 500 point locations on a real focus sequence were randomly selected and the features of the

Table 3.4: Spearman's correlation of different quality measures

	Range	Fmax	Nlev	Res	W	Q_r	E_{rms}
Range	1	-0.50*	0.79*	0.86*	0.28	0.31	**-0.31**
Fmax		1	-0.06	-0.35**	-0.25	-0.38*	**0.38**
Nlev			1	0.87*	0.43*	-0.32**	**0.32****
Res				1	0.59*	0.04	**-0.04**
W					1	-0.40**	**0.40****
Q_r						1	**-1**
E_{rms}							1

(*$p < 0.05$, ** $p < 0.1$)

focus functions corresponding to those positions computed, namely: the range (Range), width (W), number of false maxima (Fmax), the resolution (Res) and the noise level (Nlev); as well as the proposed relative quality measure, (Q_r), and the reconstruction error (E_{rms}). Table 3.4 shows the Spearman's rank correlation coefficients among those variables.

From table 3.4 it is clear that the quality measure Q_r is strongly correlated to the quality of the reconstruction in terms of the reconstruction error, with the advantage of being independent of the units and particular characteristics of each experiment. In contrast, the quality measures derived from features of the focus function are redundant and not as strongly correlated with the reconstruction error. In fact, only the noise level (Nlev) and the width (W) of the focus function showed a positive correlation with the reconstruction error above the significance level (p-value $p < 10\%$). Based on these results and, for the sake of brevity, in the sequel, the relative quality, Q_r, is preferred for analyzing the performance of the focus measure operators.

As for the reconstruction accuracy, when the focus measure operators are applied to different image sequences within the same group (e.g., focus sequences from the same acquisition device), similar rankings were obtained with respect to their quality measures, both E_{rms} and Q_r. For instance, Fig. 3.4 shows the mean relative quality of each focus measure operator on sequences from different acquisition devices. Each color bar corresponds to the mean performance for four sequences from the same acquisition device (thus, there are three bars for each focus measure operator). This figure shows that, although the overall behavior is similar among different acquisition devices, there are small differences among the obtained rankings. Thus, a slightly different ordering is obtained if the operators are sorted according to their performance on one particular acquisition device. This is reasonable if the characteristics of each device are considered: on the one hand, the sequences from the synthetic set represent an ideal case, without noise or optical artifacts. On the opposite side, the sequences obtained with the webcam have the highest effects of noise, radial distortion, image field curvature and vignetting due to the quality of the camera's optics. The difference in quality,

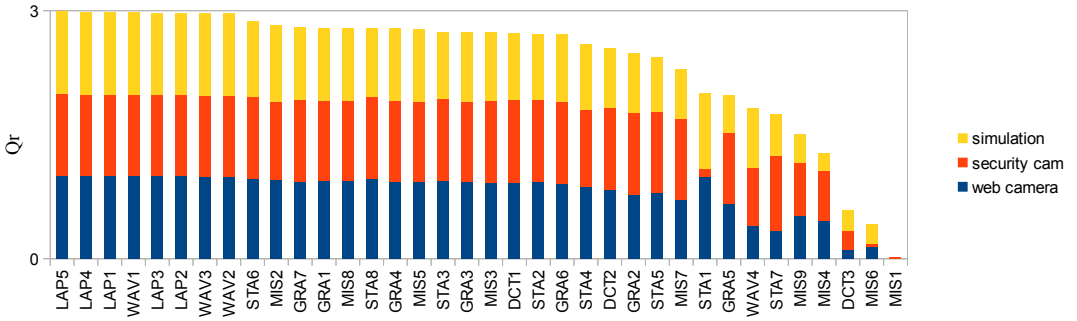


Figure 3.4: Mean performance of focus measure operators according to the relative quality Q_r (vertical axis) for different groups of focus sequences. Each color corresponds to focus sequences acquired with the same acquisition device.

content and nature of the acquired images can favor some operators while being detrimental to others according to their sensitivity to these factors. In Fig. 3.4, the focus measure operators have been sorted according to the overall Q_r obtained for all the focus sequences for displaying purposes. The rankings obtained from different acquisition devices show correlation coefficients between 0.70 and 0.93 with a p-value $p < 0.01$.

The results presented in Fig. 3.4 show that the overall ranking of focus measure operators is related to the imaging device and the scene. Thus, it is difficult to determine with certainty what operator or group of operators will perform better under particular imaging conditions. As stated in the previous paragraph, this can be explained by considering that each combination of variables, such as the imaging device and the real scene under observation, represent a different scenario in terms of contrast, noise, saturation, etc. Therefore, a given focus measure operator will perform worse or better than others according to its sensitivity to the aforementioned factors. In addition, as shown in the next section, the size of the evaluation window also affects the relative performance of focus measure operators. Notwithstanding, some global trends can be observed, such as some operators that generally exhibit a good performance in all cases (e.g., Laplacian-based and wavelet-based operators) or others that yield the worst performance in general (e.g., MIS1, MIS6).

At this point, the conducted experiments suggest that the family of Laplacian-based operators have the best overall performance at *normal* imaging conditions (i.e., without the addition of noise, contrast reduction or image saturation). The image Laplacian is a discrete approximation of the second derivative of the image and highlights rapid changes in intensities. This makes it suitable for detecting changes in focus. These results are in agreement with Russell and Douglas (2007) for autofocus applications, and Subbarao and Tian (1998) for shape-from-focus. In contrast, Sun et al. (2004) found that statistics-based methods have a better

performance in autofocus.

3.5.2 Response to imaging factors

In order to assess the robustness of the evaluated focus measure operators to image noise, contrast, saturation and size of the support window, several operators were pre-selected based on their overall performance. In this section, the performance of the operators is evaluated by taking into account their working principle. Thus, only the three best operators of four families were considered: Laplacian-based operators (LAP5, LAP4 and LAP1), wavelet-based operators (WAV1, WAV3 and WAV2), image statistics-based operators (STA6, STA8 and STA3) and gradient-based operators (GRA7, GRA1 and GRA4). Miscellaneous operators were not considered since they are based on different working principles and DCT-based operators were not included due to their low performance and high computational times according to the previous experiments. The aim of pre-selecting a set of focus measure operators and grouping them according to their working principles is to facilitate the discussion and interpretation of the results. As shown in the following sections, operators based on similar concepts exhibit a comparable response to changes in image conditions.

Sensitivity to support window

In general, the performance of all focus measure operators decreases for small evaluation windows. Fig. 3.5a shows the mean E_{rms} obtained by the different operators for all the real sequences of the test set. From this figure it is evident that, in general, E_{rms} increases as the support window size is reduced. This is expected since the size of the evaluation window directly affects the amount of texture and image information that makes it possible to detect changes in focus. However, as already noted by previous researchers, increments in the window size yield a reduction of spatial resolution[3]. Malik and Choi (2007) pointed out that increasing the window size yields a reduction of the quality of the reconstruction by excessively smoothing the depth-map. Therefore, the optimum window size for a particular application must be a trade-off between spatial resolution and robustness to the lack of texture.

In order to compare the influence of the window size, Fig. 3.5b shows the mean of the relative quality measure Q_r for all the analyzed focus measure operators. The results indicate that the differences between the Laplacian-based and statistics-based operators tend to decrease as the size of the evaluation window is

[3]In this work, the reduction of spatial resolution did not lead to greater errors probably since the reconstructed scenes mostly consist of planar patches.

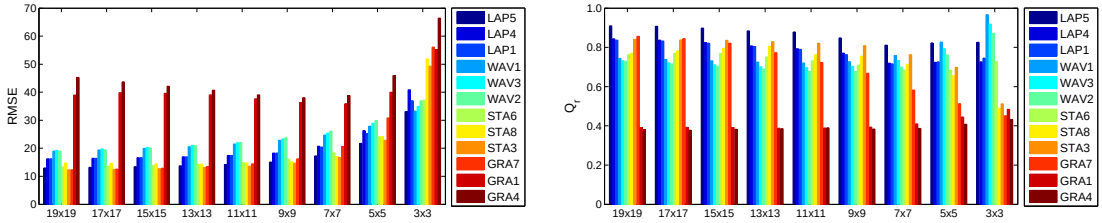


Figure 3.5: Sensitivity to window size. Mean E_{rms} in mm/pixel for all real sequences (left) and relative quality for all the real sequences (right). x-axis: window size.

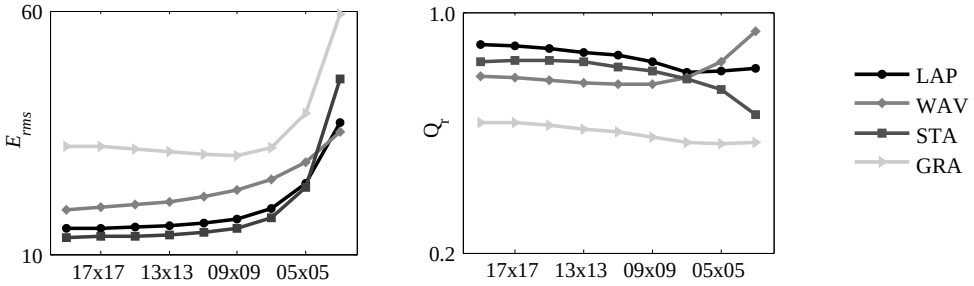


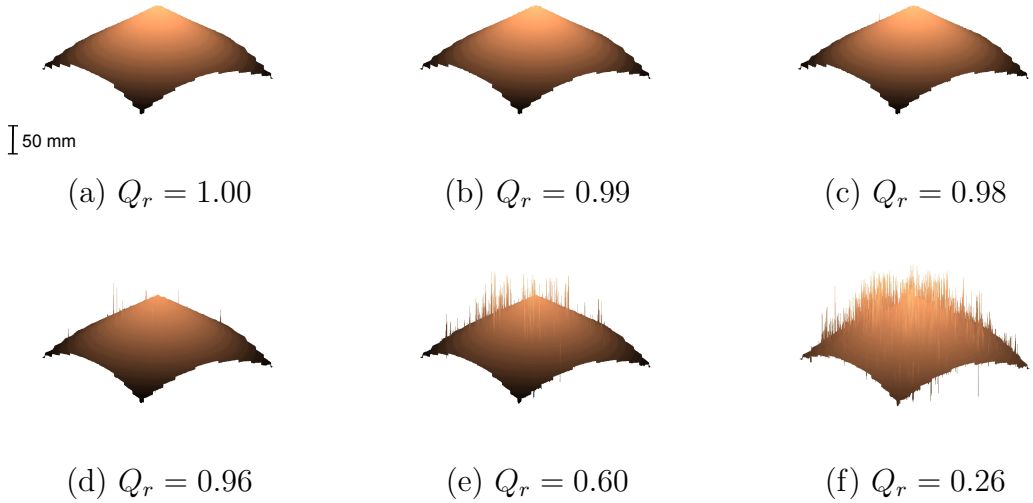
Figure 3.6: Performance for different families of operators and different window sizes. E_{rms} in pixels/mm (left) and relative quality measure (right). x-axis: support window size.

reduced. On the other hand, the gradient-based operator, GRA7, is the operator most affected by the reduction of window size. In contrast, the performance of the wavelet-based operators shows a significant improvement for small support windows.

From the tests described in this section and throughout this work, it is possible to observe that focus measure operators based on similar concepts respond similarly to variations in the imaging conditions. Therefore, for the sake of clarity, it is easier and more meaningful to understand the behavior of the various families of focus measure operators instead of each operator on its own.[4] In this way, Fig. 3.6 shows the mean performance of each family of focus measure operators after averaging the quality measures obtained by the operators within the same family for each window size.

Fig. 3.6 confirms that wavelet-based operators perform better for small evaluation windows, whereas gradient-based operators are the most sensitive to this feature. As stated before, the wavelet decomposition of an image can be interpreted as a simultaneous frequency and scale-space analysis where the detail sub-bands

[4]Individualized detailed performance of the pre-selected focus measure operators can be found at http://www.sayonics.com/research/focus_measure.html



(a) $Q_r = 1.00$ (b) $Q_r = 0.99$ (c) $Q_r = 0.98$

(d) $Q_r = 0.96$ (e) $Q_r = 0.60$ (f) $Q_r = 0.26$

Figure 3.7: Depth-maps obtained with SFF using WAV1 (top row) and LAP2 (bottom row) for different window sizes (7×7, 5×5 and 3×3, from left to right). These depth-maps correspond to the left-most synthetic sequence of Fig. 3.3. Z-axis: pixel depth (mm).

are related to the highest frequencies of the image (Gopinath et al., 1994). In addition, according to the theory of defocus, changes in focus mostly affect the high frequency components of the image. This explains why wavelet-based operators improve their relative performance as the window size decreases. Actually for small windows, the change in focus can successfully be detected at the coefficients of the low scale sub-bands of the DWT. For illustration, Fig. 3.7 compares the depth-maps obtained using a wavelet-based operator (WAV1) against those generated with a Laplacian-based operator (LAP2) for different window sizes. It is evident that their performance is comparable for the largest window. However, as the size decreases, the response of the Laplacian-based operator quickly deteriorates, while the wavelet-based operator responds more robustly. The sequence used to generate the depth-maps of Fig. 3.7 and Fig. 3.9 corresponds to the left-most synthetic sequence of Fig. 3.3.

It is also important to remark that, for a certain evaluation window size, the ranking (i.e., the quality measure) of operators may vary depending upon the image set used to perform the tests. Thus, the ranking for a given window size will differ if the sequences from the surveillance camera, the webcam or both are used. The curves in Fig. 3.6 correspond to the average performance on all the real sequences of the test set. The individual curves corresponding to different acquisition devices showed a minimum correlation of 0.81 for a p-value $p < 0.01$. The high correlation between sequences from different sources guarantees that the overall behavior of

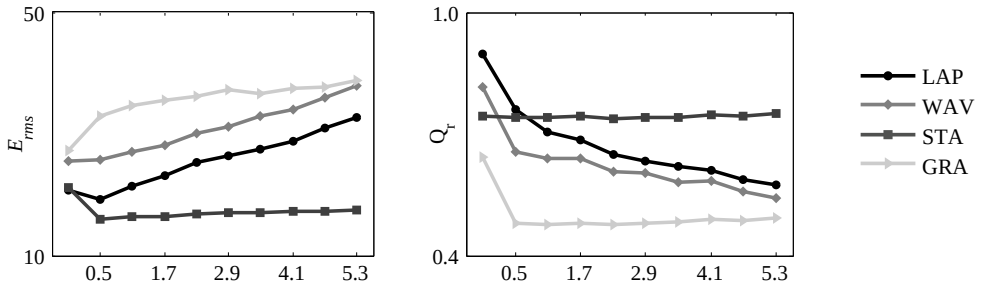


Figure 3.8: Average performance for different families of operators and noise levels. E_{rms} in mm /pixels (left) and relative quality (right). Minimum correlation of individual sequences: $r = 0.97$ for $p < 0.01$. X-axis: noise variance $\sigma_c = \sigma_s$ ($\times 10e^{-3}$).

the different operators is independent of the image sequences used to perform the tests and the same tendencies hold independently of the acquisition device. This reasoning is followed for the tests described in the next sections. Thus, only the average results along with the minimum correlation corresponding to sequences from all the sequences will be shown.

Robustness to noise

Following the procedure described in section 3.4.3, the robustness to noise of the focus measure operators has been assessed by performing 3D reconstructions under ten different noise levels. The results are summarized in Fig. 3.8. In these experiments, image noise was one of the factors that most affected the performance of all operators, with all the measured E_{rms} increasing with the amount of noise level. From this figure, it can be realized that statistics-based operators have the highest robustness to noise, with the STA2 operator being the best. On the other hand, Laplacian-based operators, which exhibit the best performance at the lowest noise level, are highly sensitive to noise, showing the greatest reduction in their relative quality measure.

The sensitivity to image noise of Laplacian-based operators is a well known fact and the robustness of statistics-based operators is in agreement with the theoretical working principles presented in section 3.2. On the one hand, low-order statistical moments are theoretically expected to have a low correlation with the high-frequency components of noise. For instance, according to the properties of linear functions of random variables, variance-based operators will detect focus accurately provided that the variance of noise is below the one of the signal (intensity values). This can be readily demonstrated as follows.

Let us express the noisy image in (3.4) as:

$$I_n = I_0 + n + n[I_0], \tag{3.7}$$

where I_0 is the ideal noiseless image, $n[I_0]$ an image-dependent noise component and n and image-independent noise component.

By considering the image (and noise components) as random variables, the variance of the noisy image can be estimated as a function of the variances of each component in (3.7) and the covariance between the source radiance and the image-dependent component, $Cov(I_0, n[I_0])$ (Montgomery and Runger, 2010):

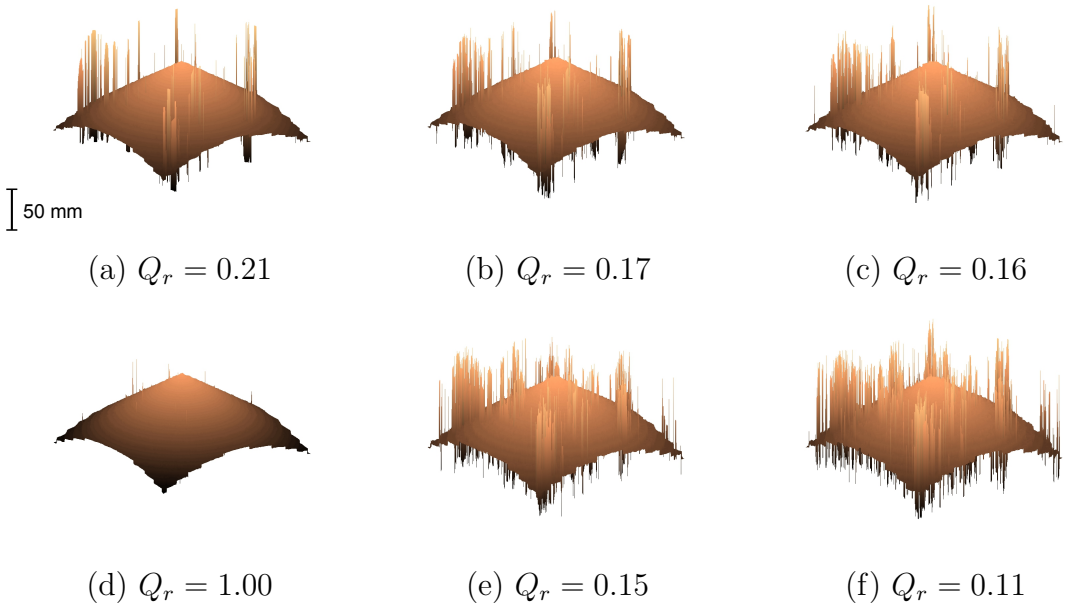
$$Var(I) = Var(I_0) + Var(n) + Cov(I_0, n[I_0]) \cdot Var(n[I_0]) \tag{3.8}$$

In the literature of image denoising (more precisely in image restoration), the correlation between the source irradiance and the image-dependent noise component is often neglected, such that $Cov(I_0, n[I_0]) \to 0$. As a result, the variance of the noisy image can be considered as an accurate estimator of the ideal noiseless image as long as $Var(I_0) > Var(n)$. This explains the robustness of statistics-based operators to image noise. In contrast, Laplacian-based operators are the most sensitive to noise since it is well known that second derivatives are very sensitive to it (Juneja and Sandhu, 2009). For illustration, Fig. 3.9 compares the depth-maps obtained using a statistical-based operator (STA3) against those generated with a Laplacian-based operator (LAP5). It is evident that the Laplacian-based operator has a better performance for the lowest noise level. However, as the noise level increases, its performance quickly deteriorates with respect to the statistical-based operator.

In addition, as shown in Fig. 3.8b, wavelet-based operators also have a high sensitivity to noise. This can be explained by the fact that image noise mostly corresponds to high-frequency components. Therefore, it is likely to have an impact on the coefficients of the detail sub-bands of the DWT. This is not surprising since, in the literature related to image denoising, noise is indeed often suppressed by thresholding the coefficients of the DWT sub-bands (Chang et al., 2000; Portilla et al., 2003). Thus, the wavelet-based focus measure will deteriorate due to the effects of noise on these sub-bands.

Sensitivity to image contrast

The experiments in this section show that the effects of contrast on the performance of focus measure operators are marginal, only with a slight increase in the E_{rms} of the obtained depth-maps, even for contrast levels reduced up to 10%. Moreover, the relative performance of the focus measure operators remains almost unaltered for the different contrast levels. The results of the evaluation of focus measure operators for different contrasts are summarized in Fig 3.10. From them, it can



(a) $Q_r = 0.21$ (b) $Q_r = 0.17$ (c) $Q_r = 0.16$

(d) $Q_r = 1.00$ (e) $Q_r = 0.15$ (f) $Q_r = 0.11$

Figure 3.9: Depth-maps obtained with SFF using STA3 (top row) and LAP5 (bottom row) with a 7×7 window and increasing noise levels ($\sigma_{s,c} = 0, 1.1e{-}3$ and $2.3e{-}3$). These depth-maps correspond to the sequence shown in Fig. 3.3i. Z-axis: pixel depth (mm).

be concluded that contrast affects all the compared operators similarly, since their relative performance almost remains unchanged. According to this figure, gradient-based operators showed the highest increase in E_{rms}.

It is important to remark that these results should be interpreted carefully. In real conditions, a reduced contrast is usually accompanied by a reduced signal-to-noise ratio, since noise in the digital image is mostly independent of the source irradiance, as previously stated. In turn, the contrast compression achieved by means of (3.5) reduces the strength of both the signal and the noise proportionally. The results in this section are aimed at assessing the effect of contrast independently of other imaging factors. This analysis also applies to the next imaging factor: the image saturation. A robust methodology for automatically assessing the effect of the image content in the estimation of the focus measure in realistic conditions will be presented in chapter 5 based on a novel theoretical focus model.

Sensitivity to image saturation

As can be observed in Fig. 3.11b, the performance of focus measure operators remains unaltered for saturations below 30%. In general, all operators decrease their performance as the saturation level is high, but this behavior is more evident for saturation levels above 30%. This can be explained by the fact that, for the

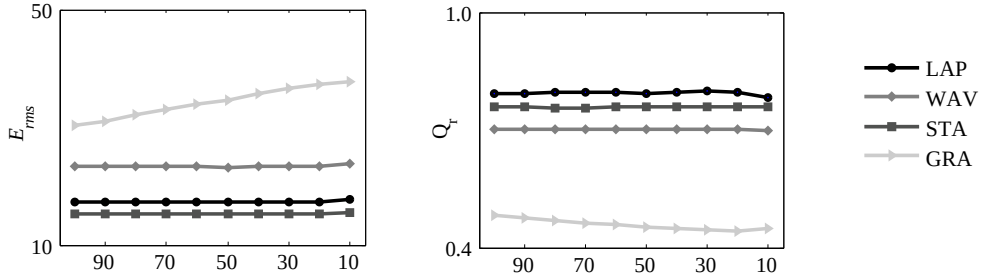


Figure 3.10: Average performance for different families of operators and contrast levels. E_{rms} in mm/pixel (left) and relative quality (right). Minimum correlation of individual sequences: $r = 0.97$ for $p < 0.01$. x-axis: contrast level (%).

imaging conditions of the captured sequences, the upper bounds of the image histogram only contain a small percentage of the total energy. Therefore, low saturation levels only affect a small fraction of image pixels. Thus, the effect of saturation is only significant above 30%. Above this threshold, the relative quality Q_r indicates that Laplacian-based operators are slightly more sensitive to saturation, as can be appreciated in Fig. 3.11b.

In general, features such as image contrast and saturation affect all the analyzed operators similarly. Thus, none of the operators is significantly more sensitive to this factor than the others and a small difference in the relative performance of some operators can only be observed at high levels of contrast and the increment of saturation. Both, the reduction in image contrast and the increment of saturation can be thought of as a reduction in the pixel depth or, in terms of image intensities, as a reduction in the number of gray levels. In this work, all the focus measure operators have been implemented in double-precision arithmetic. Therefore, no quantization or overflow problems arise in the computations independently of the operations performed by each focus measure operator.

3.6 Summary

Focus measure operators are a fundamental part of 3D scene reconstruction through shape-from-focus and shape-from-defocus, autofocus and image enhancement through focus stacking. In this chapter, a methodology to compare the performance of several focus measure operators has been proposed and tested. The selected operators have been chosen from an extensive review of up-to-date literature. Since some of them were originally proposed for autofocus applications, it has been necessary to adapt them in order to be applicable to depth estimation via shape-from-focus. Experiments have been carried out on a test set constituted by both synthetic and

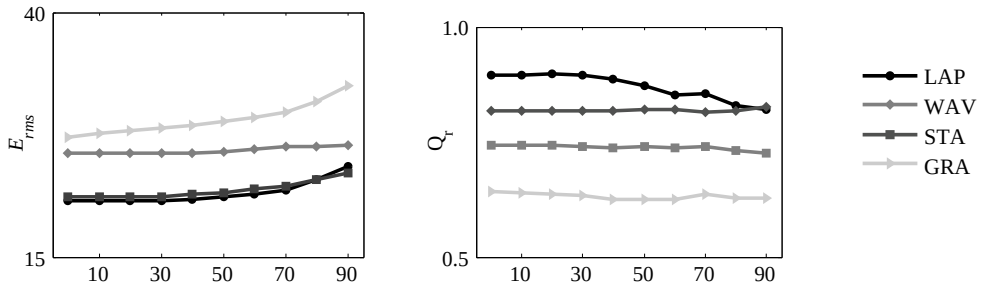


Figure 3.11: Quality measures for different families of operators and saturation levels. E_{rms} in mm/pixel (left) and relative quality (right). Minimum correlation of individual sequences: $r = 0.70$ for $p < 0.03$. x-axis: saturation level (%).

real image sequences.

The analyzed focus measure operators are based on different mathematical principles. From an initial group of 36 operators, the best 12 operators were chosen in order to compare their performance under different imaging conditions. The selected group includes algorithms based on the image Laplacian, image statistics, the image gradient and the wavelet transform.

Experiments have shown that Laplacian-based operators have the best overall performance at normal imaging conditions (i.e., without addition of noise, contrast reduction or image saturation). However, it is difficult to determine which focus measure operators have the best performance for specific imaging conditions (i.e., a given noise level, contrast, saturation and window size), since this strongly depends on the particular capturing device with which the image sequences are acquired. As a result, obtaining an absolute ranking of individual focus measure operators is rather an unfeasible task. This suggests that the results obtained by previous researchers aimed at ranking different focus measure operators according to their performance should be utilized with care since these results can be highly biased by the selected test sets. Interestingly enough, the overall behavior of the different operators is related to their working principle independently of the capturing device and they respond similarly to changes in noise, contrast, saturation and window size even for different devices. This conclusion is sustained in the fact that, in the experiments presented in section 2.3, the relative quality measures corresponding to different sequences were highly correlated beyond the significant statistical level. Experiments have also shown that operators belonging to the same family, which are thus based on similar principles, have a similar response to changes in the imaging conditions.

In summary, the results presented in this dissertation provide an insight on how different imaging conditions can affect the different families of focus measure op-

erators. Moreover, the group of best operators for SFF has been identified, which can be useful for future development of new focus measure operators and reconstruction schemes in this particular field. In particular, Laplacian-based operators showed a good overall performance at normal imaging conditions (without the addition of noise or modifications to the image contrast or saturation). Wavelet-based operators showed an improved relative performance for small support windows. As for the image noise, image statistics-based operators showed the highest robustness to this factor, whereas gradient-based and Laplacian-based operators were the most sensitive ones. All these results are in agreement with previous theory in different computer vision and image processing applications, such as image denoising, image enhancement and compression, as discussed in previous sections. No significant difference on the performance of different focus measure operators was found regarding the image contrast and saturation.

The texture content is an important factor that influences the performance of SFF, as well as other focus-related applications (Sundaram and Nayar, 1997; Muhammad et al., 2009; Gaganov and Ignatenko, 2009). However, the problem of identifying what texture families and what texture features are relevant for focus detection still needs to be assessed. In the following chapters, a methodology to predict the effect of the parameters of the acquisition device on the focus measure is presented (chapter 4). In addition, the image content problem (i.e., the poor response of the focus measure operators to low-textured images) is tackled in chapters 4 and 5.

In previous research, the performance of focus measure operators for autofocus was analyzed in terms of some features of the focus function (e.g., its sharpness, width and number of false maxima (Sun et al., 2004)). However, the results in section 3.5.1 empirically show that these features are, at the best of cases, weak predictors of the performance of focus measure operators for the pixel-wise estimation of the focus level. In chapter 4, it will be shown that, in fact, some features of the focus function, such as its width, smoothness and sharpness are more influenced by the parameters of the acquisition device and the scene geometry than by the particular focus measure operator.

CHAPTER 4

Focus modeling

A widely known practical defocus model is the one based on the so-called *thin-lens* model. This model corresponds to a first-order approximation of defocus according to geometrical optics. By extending the thin-lens model, this chapter introduces a new interpretation of the concept of *depth-of-field* (DOF). Instead of the traditional formulation for the near and far limits of the DOF, the blur width measured in pixels is used to describe the behavior of the focus in conventional cameras. With this alternative formulation, a new efficient and robust calibration method is derived. In addition, a new theoretical model for the the variation of the focus level as a function of the focus of the camera (the *focus profile*) is presented. The concepts introduced in this chapter are exploited for practical focus calibration, efficient focus sampling and for predicting the reliability of the measured focus function in real imaging conditions.

Section 4.1 reviews some basic concepts and relevant previous work. The proposed calibration method and the new focus profile model are presented in sections 4.2 and 4.3, respectively. In section 4.4, the proposed theoretical focus profile is then exploited for describing the influence of the parameters of the acquisition device on the behavior of the focus function in conventional cameras. Section 4.5 presents some experiments that validate and illustrate the properties and limitations of the proposed model. The obtained results are summarized in section 4.6.

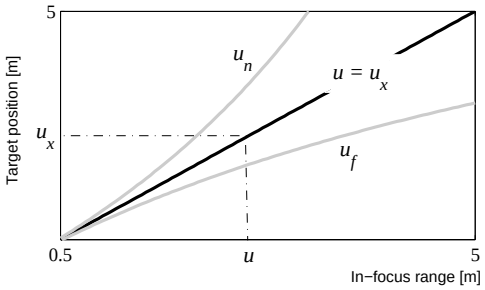


Figure 4.1: Limits of the depth-of-field. When the focus of the camera, u, points to a given target position, u_x, all the objects between the curves u_n and u_f are within the depth-of-field. Curves generated for a camera with $f = 35$ mm, $N = 4.0$ and a blurring circle of $\rho_{\max} = 35\mu$m.

4.1 Introduction

A camera with a limited depth-of-field implies that, for a particular focus setting, only the objects within a certain distance range from the camera are in focus (chapter 2). Many focus-related applications must deal with this limitation in order to optimize the amount of information captured and processed during the image acquisition process. For instance, in autofocus, it is necessary to minimize the focusing time by reducing the number of focus samples, that is, the positions where the focus needs to be measured. In the case of focus stacking, the number of images of the focus sweep must be reduced in order to speed up the acquisition process as well as to reduce the amount of memory and computational power required to perform the image fusion[1].

In this dissertation, the process of capturing images of a given scene at particular focus settings will be referred to as *focus sampling*. In some applications, such as autofocus, focus stacking and shape-from-focus, the number of focus samples should be kept to a minimum for the sake of efficiency. In contrast, too few focus samples will yield low-quality depth-maps in shape-from-focus, degraded images in focus stacking or a failure in the determination of the best focus position in autofocus. The boundaries that delimit the DOF are referred to as near limit, u_n and far limit, u_f. As stated in section 2.2, based on the thin-lens model, u_n and u_f are expressed as a function of internal parameters of the lens-camera system, such as the focus of the camera, u, the focal length, f, the f-number, N, and a predefined parameter, namely the maximum allowed blurring circle, $\rho_{\max}$. The problem of efficient focus sampling is simplified when the DOF limits are accurately known. For illustration, Fig. 4.1 plots the near and far limits of the depth-of-field as a function of the focus of the camera.

[1] At this point, the reader may be interested in revisiting section 2.2 in chapter 2 for reviewing some basic concepts and terms related to focus stacking and autofocus.

In Fig. 4.1, it is straightforward to find a set of focus positions, $\{u_k \,|\, k = 1, 2, \cdots, K\}$, which cover the whole focusing range efficiently, since the in-focus range corresponds to the vertical distance between the u_n and u_f curves for a given focus position. Unfortunately from a practical perspective, there are two drawbacks with this definition of depth-of-field and its limits. First, the assumption of known physical parameters is often difficult to accomplish in many conventional off-the-shelf cameras. As a result, the estimation of the near and far limits of the depth-of-field is not reliable. As will be shown in section 4.5, an approximated knowledge of the parameters of the acquisition device yields an inaccurate estimation of the effect of defocus. Secondly, the definition of the maximum blurring circle, or *circle of confusion*, is often empirical and requires prior knowledge about physical parameters of the imaging system and viewing conditions. Traditionally in photography, a standard circle of confusion between 0.025 mm and 0.035 mm is utilized based on some assumptions about the final printed image format, viewing distance and human visual acuity. Naturally, the selected value of the circle of confusion, and hence the focus sampling, will depend on these *ad hoc* variables.

Previous work

The problem of accurately knowing the parameters of the acquisition device is often overcome by means of calibration. Different calibration methods were reviewed in section 2.2.6. The main drawback of these procedures is the need for an extensive experimental setup in order to achieve calibration. In the best case of using a calibration Gauge, the calibration process needs to be performed at least once for each focus setting of the camera. This represents an important practical limitation in conventional cameras since, even the simplest cameras with focus capability, such as cellphone cameras, can be set to a large number of different focus positions.

In order to overcome this problem, Muhammad and Choi (2012) developed a focus sampling criterion for shape-from-focus in microscopy. In that work, the optimum sampling frequency F_s (the number of focus samples as a function of the distance from the camera) is derived in the frequency domain using the Nyquist criterion as $F_s = DOF/4\alpha$, where $\alpha \approx 3$ is a constant. As stated before, for conventional cameras with known parameters, the DOF can be estimated from the near and far limits as $DOF = u_f - u_n$. In optical microscopy, the DOF is estimated from the refractive index of the media, the light's wavelength and the numerical aperture of the microscopy imaging system. Unfortunately, the focusing process in microscopy has important differences with respect to conventional cameras. As a result, the extension of that approach to this type of cameras is limited.

Vaquero et al. (2011) adjusted the speed of the focus mechanism of a mobile camera as a function of the near and far limits of the DOF and the shutter speed. Although u_n and u_f can only be approximately computed as a function of the

parameters of the camera, that method can still be applied for an efficient focus sweep whenever the speed of the focus mechanism can be controlled accurately. The problem of time-efficient focus sampling was studied in detail by Hasinoff and Kutulakos (2011) for cameras with an accurate control and explicit knowledge of the optical parameters (namely, exposure time, focal length, aperture, and speed of the focus mechanism). Alternatively, the calibration procedure that will be presented in section 4.2 implicitly obtains parameters of the acquisition device, allowing for an accurate description of the DOF limits without any explicit knowledge about the real physical parameters of the camera. The focus-calibrated camera can be exploited for efficient focus sampling for different applications, such as autofocus (section 4.5), focus stacking and shape-from-focus (chapter 5).

In photography, the formulation for the near and far limits of the DOF based on the thin-lens model allow an approximate determination of u_n and u_f (equations 2.19 and 2.20). Thus, manufacturers generate DOF look-up tables that depend on the camera model, more specifically, on the sensor size and the chosen circle of confusion [2]. The photographer can then use those values as a reference, being able to manually adjust focus in order to obtain an accurate acquisition. The calibration method proposed in this chapter aims at implicitly estimating the variables that define the DOF in order to allow an accurate and *automated* focus sampling.

4.2 New calibration and sampling methods

In order to derive a calibration method for efficient focus sampling, let us divide the problem in three parts: in the first one, a new expression for describing the amount of defocus as a function of the lens parameters and the focus of the camera is derived based on the thin-lens model. Secondly, a method for practical implicit calibration is proposed. Finally, an efficient focus sampling methodology is presented.

4.2.1 Defocus blur

A useful and widely known way of assessing the effect of defocus is the *blur radius*, ρ, is derived from the thin-lens model. For convenience, let us rewrite equation (2.4) as:

$$\rho = v\frac{D}{2}\left(\frac{1}{f} - \frac{1}{u_x} - \frac{1}{v}\right), \tag{4.1}$$

where D is the lens diameter, f the lens focal length, u_x the target position and v the internal position of the sensor with respect to the lens (Fig. 2.5).

[2] See for instance www.dofmaster.com/dofjs.html

Although (4.1) is a simple closed-form expression that expresses the defocus blur as a function of the camera parameters, it has three important limitations that prevent the formulation of a practical calibration procedure:

1. The blur radius is a function of the internal lens configuration, v. The distance between the sensor and the lens is difficult to be determined experimentally. In most cases, this variable is simply a parameter of the model with no real physical meaning (e.g., in compound lens systems). Alternatively, it would be desirable to describe ρ as a function of the in-focus distance, u, which is an external real physical parameter of any camera with focus control.

2. The blur radius is measured in metric units. The units of the blurring circle in (4.1) depend on the units of the physical parameters of the camera (millimeter, meters, etc). Instead, from an image processing perspective, it would be desirable to measure the blur width in pixels. As an example, consider the case when, in a particular experiment, an object is said to be imaged with a blurring radius of $15\mu m$. Without additional knowledge about the resolution of the system, it is not possible to determine whether this amount of blur is negligible or not. In contrast, if the same object is said to be imaged with a blurring circle of 0.5 pixels, it could be reasonably assumed that the blur is negligible. In some particular applications, the metric units are preferred, for example, when comparing the resolution or quality of two different acquisition devices.

3. The blur radius is a function of the parameters of the camera. As a result, when *any* of the parameters in (4.1) changes, so does the blur width. Alternatively, it would be desirable to analyze the blur width only as a function of both the target position u_x and the focus of the camera, u. This would allow a practical calibration by changing the focus of the camera instead of keeping a fixed focus and moving the target (as in calibration procedures proposed so far). Again, the statement of the problem depends on the application: in this particular case we are interested in the effects of focus. In contrast, other applications could be interested in the effects of zoom or the lens aperture and the problem should be re-stated accordingly.

In order to tackle the limitations listed above, starting from (4.1) and after some simple algebraic manipulation, the blur width can be computed as:

$$\rho = \frac{1}{N}\left[(v - f) - \frac{f}{u_x}v\right], \tag{4.2}$$

where $N = f/D$ is the f-number of the camera. In the sequel, the f-number has been preferred over the lens diameter since it is the typical control found in conventional cameras.

As previously stated in section 2.2.6, according to the paraxial approximation of geometrical optics, the relationship between the sensor position and the focus of the camera is governed by the thin-lens or *lens maker's* equation:

$$\frac{1}{f} = \frac{1}{v} + \frac{1}{u} \tag{4.3}$$

The lens maker's equation is valid for single-lens systems but has shown to be quite useful for describing the relationship between the image plane and the focus position for complex imaging systems, including compound lenses (Horn, 1990). From this equation, it can be readily shown that:

$$v = \frac{uf}{u - f} \tag{4.4}$$

The internal parameter v can be removed from (4.2) by replacing it according to (4.4), yielding:

$$\rho = \frac{f^2}{N} \frac{u_x - u}{u_x(u - f)} \tag{4.5}$$

Equation (4.5) is valid for $u_x > u$. With respect to Fig. 2.5 and using similar triangles, it is straightforward to verify that this equation is also valid for $u_x < u$ by simply taking the absolute value. In addition, in order to be able to measure the blur width in pixels, an additional parameter must be introduced: the pixel density of the sensor, γ. The pixel density measures the amount of pixels of the sensor per metric unit (e.g., in pixels per millimeter). At this point, the introduction of yet another parameter of the camera could seem counterproductive. However, the benefits will be evident in the next sections. By scaling (4.5) by the pixel density, the blur radius (measured in pixels) can be computed as:

$$\rho = \frac{\gamma f^2}{N} \frac{|u - u_x|}{u_x(u - f)} \tag{4.6}$$

In order to further simplify (4.6), it is necessary to elaborate on a manufacturing detail of most conventional cameras: from (4.3) it is clear that, in order to focus objects at short distances (small u), the lens-sensor distance rapidly increases (hence $v \to \infty$ as $u \to f$)[3]. As a result, most real lenses have a compulsory

[3]Notice that (4.3) has a singular point at $u = f$. Due to design limitations in real systems, only the solution for $v \geq f$ is considered.

minimum in-focus distance due to the restrictions on the physical dimensions of the system (Allen and Traintaphillidou, 2011). This restriction increases for large focal lengths. For instance, a *normal* lens AF-S DX NIKKOR has a minimum focus distance of 280 mm for a focal length of 18mm[4]. For the case of the human eye, this can easily be confirmed since it is not possible to focus on objects too close to the eye. Thus, (4.6) can be simplified by assuming that $u >> f$, yielding:

$$\rho \approx \kappa \frac{|u - u_x|}{u u_x}, \tag{4.7}$$

where $\kappa = \gamma f^2/N$ is a single constant that accounts for the camera parameters. In the sequel, κ will be referred to as the *camera constant*. Importantly, the camera constant is independent of the focus setting of the camera.

The practical advantage of the model in (4.7) is evident since it groups the effects of the physical parameters of the lens in a single value, the camera constant, while expressing the blur radius as a function of the target position and the focus of the camera. As a result, it is possible to find the parameter of the model, κ, by simply changing the focus of the camera, u, and measuring the blur radius ρ. Notwithstanding, a method for measuring the blur radius by processing a defocused image still needs to be formulated. This problem is tackled in the next section.

4.2.2 Calibration method

In order to be able to determine the value of the camera constant defined above, it is necessary to measure the blur radius as a function of the focus of the camera. Fortunately, this problem is analogous to the experimental estimation of the resolution of digital imaging systems. The problem of estimating the resolution of digital systems directly from the captured images has previously been tackled for radiographic systems and tomographic scanners. In that scope, Judy (1976) exploited the *edge spread function* (ESF), that is, the response of an optical system to a step edge, in order to determine the resolution of a computer tomographic scanner. Subsequently, Samei et al. (1998) extended and improved this method for its application to radiographic systems. In this analogy, a camera with changing focus can be interpreted as an optical system with variable resolution, whereas the resolving capability of the system is inversely proportional to the blur width. For example, Fig. 4.2a shows an edge target imaged with different amounts of defocus. In this figure, an increasing level of defocus can be interpreted as a reduced resolving capability of the acquisition device. Inspired by this analogy, the method

[4]Normal lenses are those with a focal length near 35mm that yield a similar perspective to the human eye.

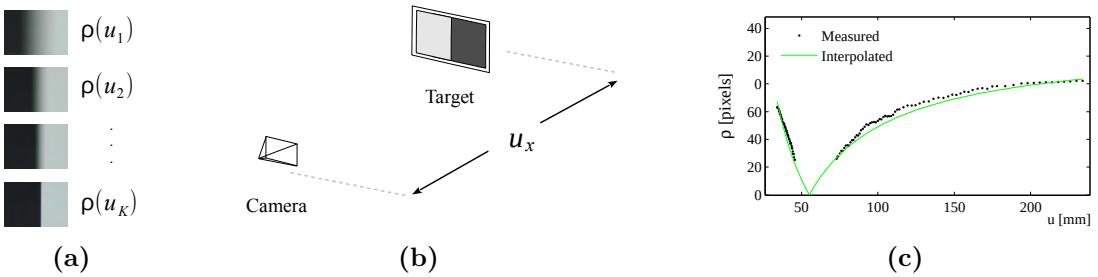


Figure 4.2: Calibration of camera focus. (a) Sequence of images of an edge target captured at different focus settings. (b) The calibration edge target should be placed at an approximately known distance u_x from the camera. (c) The camera constant is found by fitting (4.9) to the measured blur radius.

for measuring the blur width from the ESF in order to calibrate the camera is described below.

For simplicity and without loss of generality, let us consider the 1D equivalent of the Gaussian PSF of a blurred system:

$$h_\rho(r) = \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{r^2}{2\rho^2}\right), \tag{4.8}$$

where $r^2 = x^2 + y^2$.

By considering the optical system as a linear shift-invariant system (chapter 2), the response of the imaging system to a step image, $I_u(x)$, conforms to $I_D(x) = h_\rho(r) * I_u(x)$, yielding (Kayargadde and Martens, 1996):

$$I_B(x) = \frac{1}{2}\text{erf}\left(\frac{x}{\sqrt{2}\rho}\right) + \frac{1}{2}, \tag{4.9}$$

where $I_B(x)$ is the blurred observed image and erf$(\cdot)$ is the error function (Cody, 1969; Hart, 1968):

$$\text{erf}(x) = \frac{2}{\sqrt{2}} \int_0^x e^{-t^2}\, dt \tag{4.10}$$

Therefore, the blur radius can be estimated by simply fitting (4.9) to the intensity profile of a blurred edge step. Thus, the calibration setting consists of a step edge placed in front of the camera at a fixed distance u_x (Fig. 4.2b). Conveniently, the target position u_x must only be known approximately. A sequence of images is then captured by changing the focus of the camera, u, and a set of defocused edge images is generated (as the one in Fig. 4.2a). For each blurred image, I_k, of the image set, a blur width, ρ_k is computed by adjusting (4.9) to the horizontal

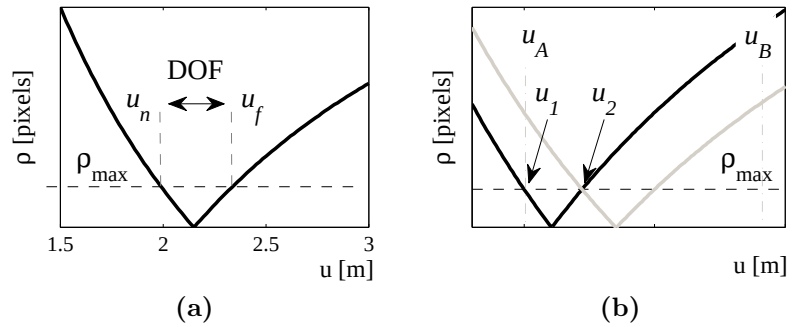


Figure 4.3: Focus sampling. (a) The near and far limits of the DOF correspond to positions where the blur radius is $\rho_{\max}$ ($\kappa = 100$ [pixels] and $u_x = 2.15$ [m]). (b) The focus samples are iteratively computed by finding the near and far limits of the DOF corresponding to the current in-focus position in order to guarantee that $\rho \leq \rho_{\max}$.

gray-level of the blurred step edges. The calibration is then completed by adjusting the curve ρ_k vs. u_k to (4.7) through non-linear optimization, as illustrated in Fig. 4.2c. This allows finding the camera parameter κ as well as refining the target position u_x.

4.2.3 Focus sampling

In the previous section, a calibration procedure that allows finding the blur width as a function of the focus of the camera was presented. For a given circle of confusion, $\rho_{\max}$, this model can be used to determine the near and far limits of the depth-of-field, as shown in Fig. 4.3a. The advantage of a blur width measured in pixels instead of in metric units is that the effect of blur can be assessed "as seen" from an image processing perspective. Thus, the maximum allowed blur is not a function of the intended printed format or viewing distance anymore, but on its effect on the digital image.

Intuitively, one could expect that a reasonable maximum blur should be below one pixel. A detailed discussion about the value of $\rho_{\max}$ will be presented at the end of this section. At this point, let us assume that the maximum blur has a known fixed value. In this case, an efficient focus sampling should guarantee that the focus samples are distributed so that the blur width of any object within the given focusing range is always below $\rho_{\max}$. In addition, the number of samples should be kept to a minimum. The proposed sampling scheme is described as follows.

Let u_A and u_B be the limits of the focusing range, with $u_A < u_B$. It is assumed that the objects of interest are within a finite distance range between u_A and u_B. The aim is to determine a set of positions $\{u_k | k = 1, 2, \cdots, K\}$, where K

is the total number of focus samples and $u_A \leq u_k \leq u_B$, $\forall k$. For any object at $u_x \in [u_A, u_B]$, there must be at least one focus sample, u_k, that guarantees $\rho(u_k) \leq \rho_{\max}$ for that object. That is, there must be at least one focus sample at which the object is imaged with a blurring below the maximum allowed blur. By default, the first focus sample corresponds to the first limit of the focusing range: $u_1 = u_A$. As illustrated in Fig. 4.3b, the condition $\rho(u_x) \leq \rho_{\max}$ holds for any object between u_1 and u_2 (i.e., between the left and right branches of the dark curve in this figure). From (4.7) it can be readily shown that $u_2 = u_1 + \Delta u_1$, where Δu_1 is the *sampling step* corresponding to the separation between the two focus samples, which is given by:

$$\Delta u_1 = \frac{2\rho_{\max} u_1^2}{\kappa - 2\rho_{\max} u_1} \tag{4.11}$$

For the next sample, u_2 is now the near limit of the DOF (the left branch of the gray curve in Fig. 4.3b). Thus, the positions of subsequent samples are iteratively computed by repeating the process described above until the whole focusing range is covered. Thus, the $(k+1)$-th focus sample is computed as:

$$u_{k+1} = u_k + \Delta_k, \tag{4.12}$$

where:

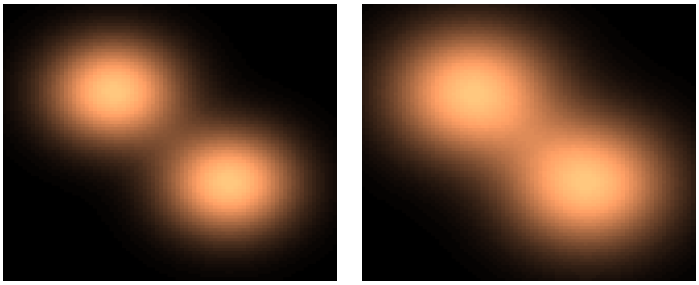
$$\Delta_k = \frac{2\rho_{\max} u_k^2}{\kappa - 2\rho_{\max} u_k} \tag{4.13}$$

The only missing parameter at this point is the maximum allowed blur, $\rho_{\max}$, measured in pixels. As it will be shown below, the advantage of measuring the blur in pixels is that it allows the definition of $\rho_{\max}$ independently of the real metric resolution of the acquisition device.

Maximum blur

The proposed focus sampling algorithm depends on two parameters: the constant κ found by calibration and the maximum allowed blur radius, $\rho_{\max}$. In the previous sections, the effect of defocus was interpreted as a variable resolution optical system. Thus, for a given focus setting, the optical resolution will depend on the degree of defocus. As a result, the maximum allowed blur should match the sensor resolution of the system in order to be able to capture the sharpest image.

In conventional cameras, the overall resolution is a function of both the sensor resolution and the optical system resolution. When the focus error is low (the camera is close to perfect focus) the resolution is said to be sensor-limited (Bass, 2010). In this case, the resolution is determined by the pixel size. As a result, the maximum blur radius should be set so that a resolution of one pixel is achieved.



(a) (b)

Figure 4.4: Maximum allowed blur as defined by applying the Rayleigh criterion. Two light spots imaged at adjacent pixels can be resolved if the FWHM of the PSF is below one pixel (a). Otherwise, the light spots cannot be resolved since the point radiances "leak" into adjacent pixels (b).

In general, it is not easy to quantify the resolution of an optical system since it depends on the signal-to-noise ratio. In practice, the FWHM (full width at half maximum) of the PSF is used as an approximated measure of the optical resolution at the Rayleigh criterion (Denton, 2000). The FWHM criterion is commonly used to determine the resolution of scanners (spot size), lasers (emission linewidth) (Bass, 2010; Marshall, 2004), telescopes and optical microscopes (Corle and Kino, 1996).

Fig. 4.4 illustrates the concept of resolution as defined by the Rayleigh criterion. In this figure, the blurred images corresponding to two infinitesimal light spots can be resolved (Fig. 4.4a) or not (Fig. 4.4b) depending on the FWHM of the respective point spread functions. In particular, the maximum blur width must be selected so that the FWHM ≤ 1 pixel. As a result:

$$2h(r, \rho_{\max})|_{r=0.5} \geq 1 \tag{4.14}$$

Solving for $\rho_{\max}$ yields:

$$\rho_{\max} \leq \frac{1}{\sqrt{8\log(2)}}. \tag{4.15}$$

If the inequality in (4.15) is satisfied, the optical resolution of the defocused system (and hence the maximum allowed blur) is as big as one pixel at most. As a result, (4.15) guarantees that the defocused lens-camera system is close to the sensor's resolution.

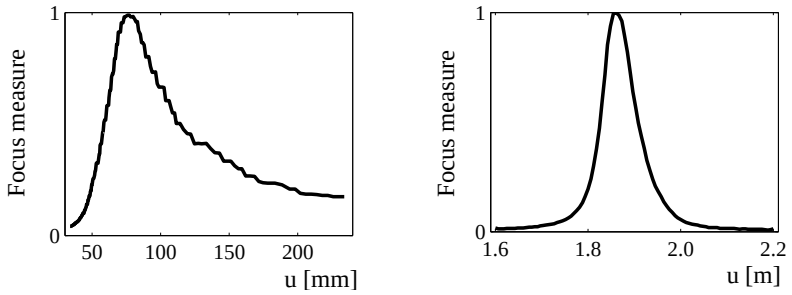


Figure 4.5: The focus profile in conventional cameras. (a) For a webcam (target at $u_x = 77$ mm). (b) For a surveillance camera (target at $u_x = 1.87$ m).

4.3 The proposed focus profile model

In this section, a new *focus profile* model for conventional cameras is introduced. From an algorithmic perspective, two fundamental concepts in focus-related applications are the focus measure and the focus profile. The focus measure has been defined and reviewed in previous chapters, whereas the focus profile simply corresponds to the normalized focus measure function, $\varphi(u)$ with $\max\{\varphi\} = 1$. Thus, in simple terms, the focus profile refers to the *shape* of the focus function, without taking into account its magnitude. For instance, Fig. 4.5 shows the focus profiles obtained from two different conventional cameras. It is clear that understanding the behavior of focus and the factors that determine the focus profile (its shape, the location of its peak, smoothness, etc) is important for different applications. Notwithstanding, few researchers have addressed the problem of theoretically modeling the focus profile in conventional cameras. To our knowledge, the state-of-the-art models are empirical (Nayar and Nakagawa, 1994) or have only been studied for high magnification systems, such as microscopes (Muhammad and Choi, 2012).

The challenge in modeling the focus profile is the high number of variables involved. The variables affecting the focus profile can be analyzed from the perspective of image processing, the scene content or the image formation process. According to image processing, the computation of the focus measure can be influenced by the image noise, contrast, saturation, the selected focus measure algorithm and its parameters (e.g., the support window). These factors were studied in detail in the previous chapter. In turn, the scene content specifically refers to the lack or presence of enough texture necessary to detect focus. In this section, a focus profile model will be proposed by taking into account mainly the factors related to the image formation process, such as the lens focal length, f-number, focus

setting, system resolution and target position. Interestingly enough, the proposed model can be exploited for dealing with the image processing factors as well as the image content problem, as will be shown in the next chapter.

Based on experimental results, Nayar and Nakagawa (1994) suggested modeling the focus profile with a Gaussian function. The Gaussian-like behavior is particularly valid near the maximum focus value and with relatively large focal lengths (Subbarao and Tian, 1998). The focus profile of Fig. 4.5b is an example of this case. The Gaussian focus profile is defined as a function of three parameters, A, μ and σ:

$$\varphi(u) = A \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) \tag{4.16}$$

More recently, Muhammad and Choi (2012) proposed to extend the model of the laser-beam propagation in order to describe the focus profile in optical microscopes. Based on this assumption, the focus profile is modeled as a Laplacian-Cauchy function:

$$\varphi(u) = \frac{A}{B + (u-C)^2}, \tag{4.17}$$

where A, B and C are the parameters of this model. These parameters, as well as those of the Gaussian profile in (4.16), are found by taking samples of the focus profile and fitting the corresponding model. Obviously, the quality of the obtained model will depend on the selected focus samples, as well as the noise and artifacts present in the measured focus profile. Independently, Tsai and Chen (2012) also proposed to exploit a focus profile as in (4.17) in order to boost the search speed in camera autofocusing.

There are mainly two drawbacks with the aforementioned two models: firstly, they have been derived for optical microscopy, which has important differences with respect to conventional cameras. For instance, focusing in optical microscopy is performed by moving the stage of the microscope while the optics remain fixed. With respect to the thin-lens model of Fig. 2.5, this would be equivalent to focusing by moving point P (instead of changing the lens-sensor distance v). Conversely, focusing in conventional cameras is achieved by changing the internal geometry of the lens-sensor system, and this affects the resulting focus profile. For instance, the asymmetry in Fig. 4.5a is due to the change of the camera's optics during focusing. This effect will be explained theoretically in section 4.4. Secondly, the parameters of the models (the constants in (4.16) and (4.17)), do not have an explicit physical interpretation. A clear relationship between the parameters of the focus profile model and the parameters of the real lens-camera system is desirable since it provides useful information about the image acquisition process.

The model proposed in this section is specifically designed to be applied to conventional cameras. Despite the number of variables involved in the focusing process, it has been derived in order to provide a simple, yet accurate, representation of the focus profile. Moreover, the parameters of the proposed model are directly related to real parameters of the lens-camera system, thus providing meaningful information about the relationship and interaction between the optics and the image-related variables involved in this process. This information may be valuable for the research community as it helps identify advantages and limitations of different focus-related applications depending on the focusing conditions, as well as to contribute to the understanding of the defocus phenomenon from both a qualitative and a quantitative perspective.

4.3.1 Theoretical focus measure

In order to determine the focus profile, the focus measure value must be expressed as a function of defocus. In chapter 2, the focus measure was defined as the energy of the acquired image after being pre-processed by means of a focus measure operator. A direct consequence of this definition is that the focus measure value is directly related to the energy of the ideal image irradiance. This interpretation is fundamental for the proposed model. Thus, in ideal conditions, the degree of focus of an image, φ, is proportional to the energy of the image itself. In the Fourier domain, by the direct application of Parseval's theorem to equation (2.12), this can be written as:

$$\varphi \approx \int\int |I(\xi,\eta)|^2 \, d\eta \, d\xi, \tag{4.18}$$

where $I(\xi,\eta)$ is the Fourier transform of $I(x,y)$ and η and ξ are the 2D frequency variables.

At this point, the problem of finding the focus profile (φ vs. u) can be solved by expressing the integral in (4.18) as a function of blur, ρ. From section 2.1.1, a blurred image, I_B, can be computed from the ideal source irradiance, I_S, and the response of the defocused system (the PSF) by means of a convolution. In the Fourier domain, this can be written as:

$$I_B(\xi,\eta) = I_S(\xi,\eta) H_\rho(\xi,\eta), \tag{4.19}$$

where $\mathcal{F}\{\cdot\}$ is the Fourier transform operator and H_ρ is the Fourier transform of the point spread function.

When H_ρ is normalized to have unity value at zero frequency, it corresponds to the so-called *optical transfer function* (OTF). According to geometrical optics,

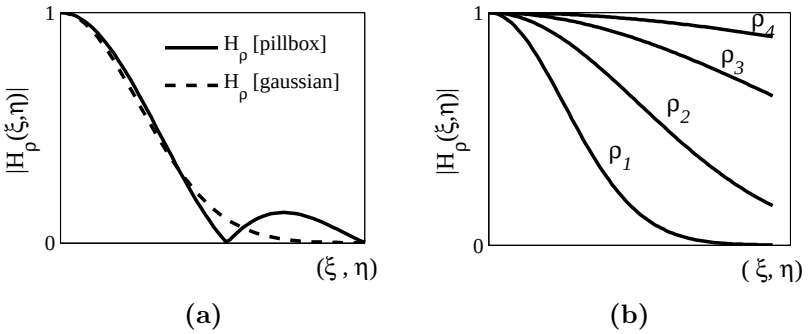


Figure 4.6: Frequency response of a defocused system. (a) Gaussian approximation. (b) Effect of blur radius, $\rho_1 > \rho_2 > \rho_3 > \rho_4$.

the OTF of a defocused optical system corresponds to the *pillbox* function. Alternatively, in incoherent polychromatic illumination, a widely used approximation corresponds to a 2D Gaussian (section 2.1.1):

$$H_\rho(\xi,\eta) = e^{-\frac{1}{2}\rho^2(\xi^2+\eta^2)} \tag{4.20}$$

Fig. 4.6a compares the cross section of the pillbox function with its Gaussian approximation. Although the two functions are not identical, in terms of energy, the Gaussian approximation is good for practical applications. The Gaussian approximation is preferred since, due to the properties of the Gaussian function and its Fourier transform in terms of separability, it simplifies the analytical development of the defocus model. In addition, as previously discussed in section 2.3.1, the Gaussian approximation has been a valid choice for multiple problems in computer vision. By analyzing the behavior of its OTF, a defocused system can be interpreted as a low-pass filtering system (section 2.1.1). For illustration, Fig. 4.6b shows the magnitude of H_ρ for different amounts of defocus. In particular, the cut-off frequency decreases with increasing blur radius. However, this defocus model is incomplete. In fact, the Gaussian approximation assumes that diffraction plays a negligible role when compared to the effect of defocus. Hence, as previously shown in section 2.3, this defocus model is not valid for small amounts of defocus. It is straightforward to verify that, near perfect focus (when $\rho \to 0$), $H_\rho(\xi,\eta)$ behaves as an all-pass system. This implies that point P is perfectly focused, which disagrees with the behavior of real systems.

In order to understand the behavior of a real system near perfect focus, it is also necessary to consider the effects of diffraction (Hopkins, 1955; Stokseth, 1969). However the equations that model this behavior are rather complex and depend on an accurate knowledge of the system's optics. In order to overcome this problem and derive a practical defocus model, an integral model of the defocused digital

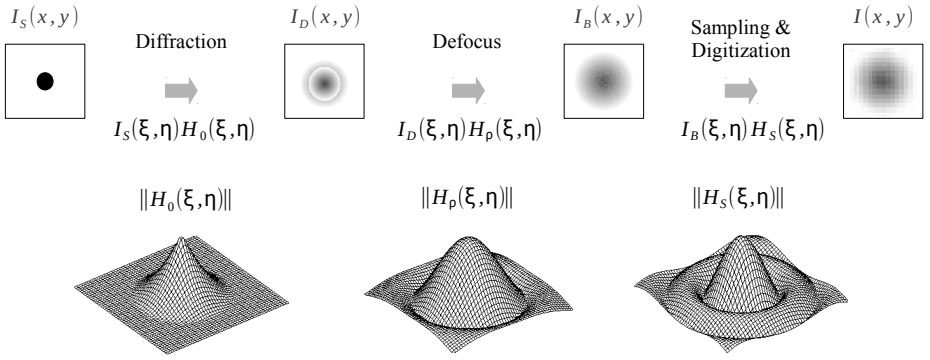


Figure 4.7: The image formation process. The digital image can be modeled as the result of cascading the effects of diffraction, $H_0(\xi,\eta)$, defocus, $H_\rho(\xi,\eta)$, and sampling, $H_S(\xi,\eta)$.

image that considers the effect of both diffraction and digitization is described below[5].

Let us consider the schematic diagram of the digital image formation process as shown in Fig. 4.7. As stated in section 2.1.1, when the wavefronts of a source irradiance $I_S(x,y)$ pass through the aperture of the imaging system, a diffraction pattern $I_D(x,y)$ is formed on the image plane (the diffraction is present even in case of perfect focus). From (2.1), this process can be modeled in the Fourier domain as $I_D(\xi,\eta) = I_S(\xi,\eta)H_0(\xi,\eta)$, where H_0 is the transfer function of the diffracting system. For a circular lens aperture, it corresponds to the Fourier transform of the Airy disc (Ersoy, 2007; Goodman, 1996):

$$H_0(\xi,\eta) = \begin{cases} \frac{2}{\pi}\left[\cos^{-1}(\frac{\omega}{c_0}) - \frac{\omega}{c_0}\sqrt{1-(\frac{\omega}{c_0})^2}\right] & \text{if } \omega < 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.21}$$

where $\omega^2 = \xi^2 + \eta^2$ and c_0 is a constant that depends on the wavelength of light and the physical dimensions of the system, as described in section 2.1.1

In Fig. 4.7, the diffracted image is then defocused in order to form the blurred image I_B. In the Fourier domain:

$$I_B(\xi,\eta) = I_S(\xi,\eta)H_0(\xi,\eta)H_\rho(\xi,\eta) \tag{4.22}$$

As stated before, in a badly defocused system, the effect of diffraction is negligible. In that case, (4.22) will be equivalent to (4.19). In the opposite case, when blur is negligible ($\rho \to 0$), H_ρ behaves as an all-pass system and, therefore, the blurred image in (4.19) is dominated by the effect of diffraction. This behavior is

[5] At this point, the reader may be interested in reviewing the relationship between diffraction and digitization given in section 2.1.2.

in agreement with the observations made by Hopkins (1955) and Stokseth (1969) about the response of a defocused optical system in presence of diffraction.

After defocus blur, the last step represented in Fig. 4.7 corresponds to the *apodization* that takes place during the digitalization of the sensed image (Ersoy, 2007). Apodization is analogous to *windowing* in digital signal processing (Oppenheim et al., 1999) and takes place when the intensity of the formed image is integrated over the pixel surface (section 2.1.2). In CCD cameras, the sampling spot is commonly a rectangular window. In this case, the Fourier transform of the apodization function (equation (2.7)) corresponds to:

$$H_S(\xi,\eta) = \frac{\sin(\xi\Delta_x/2)}{\xi\Delta_x/2}\,\frac{\sin(\eta\Delta_y/2)}{\eta\Delta_y/2}, \tag{4.23}$$

where $\Delta_x \times \Delta_y$ is the size of the rectangular window.

The digitization also implies the convolution with a train impulse and the discretization of the image, with the addition of electronic and quantization noise (section 2.1.1). For the sake of brevity, the discretization step is not detailed here since the continuous-domain representation of the imaging process suffices for the following analysis.

Finally, the whole image formation process that transforms a source irradiance I_S into a defocused digital image I can be written as:

$$I(\xi,\eta) = I_S(\xi,\eta)H_0(\xi,\eta)H_\rho(\xi,\eta)H_S(\xi,\eta) \tag{4.24}$$

According to (4.18), the focus measure conforms to the energy of the captured image. In addition, in order to simplify the notation and without loss of generality, the circularly symmetric case will be discussed in the sequel by replacing the 2D dimensional frequencies by $\omega^2 = \xi^2 + \eta^2$, yielding:

$$\varphi = \int_{-\infty}^{+\infty} |I(\omega)|^2\, d\omega \tag{4.25}$$

Replacing (4.24) in (4.25):

$$\varphi = \int_{-\infty}^{+\infty} |\tilde{H}(\omega)H_\rho(\omega)|^2\, d\omega, \tag{4.26}$$

where $\tilde{H}(\omega) = I_S(\omega)H_0(\omega)H_S(\omega)$ can be interpreted as a band-limited signal that accounts for the joint effect of diffraction, the apodization function and the high-pass filtering of the focus measure operator on the source irradiance. This substitution is introduced here for convenience in order to isolate the effect of blur, represented by $H_\rho(\omega)$, from the rest of the variables. The integral in (4.26) can be evaluated numerically in order to obtain the focus measure, φ, as a function of focus. However, a more useful closed-form solution of the focus profile can

be obtained by making some simplifications that take into account the relationship between diffraction, $H_0(\omega)$, and apodization, $H_S(\omega)$, in the image formation process.

4.3.2 Approximated focus measure

Although it is difficult to determine the particular apodization function for a sensing device *a priori*, it is well known that, in general, it behaves as a low pass-filter with cut-off frequency below the Nyquist frequency of the system. In this sense, the sampling spot works as a low-pass filter designed to prevent aliasing in the image (Pratt, 2007; Bass, 2010). Ideally, the cut-off frequencies of $H_0(\omega)$ and $H_S(\omega)$ should be close enough such that the sharpest possible image is captured.Whenever the cut-off frequency of the system is limited by $H_0(\omega)$, the system is said to be diffraction-limited. Otherwise, the system is said to be detector-limited (Bass, 2010). In either case the joint effect is, in practice, a low-pass response with a cut-off frequency ω_c that only depends on the imaging system and must be below the Nyquist frequency (Hornberg, 2006).

At this point, let us consider the source irradiance $I_S(x, y)$ as a wideband signal, such as in scenes with rich texture content, so that the spectrum of $\tilde{H}(\omega)$ is limited by the system's diffraction and apodization: $\tilde{H}(\omega) \approx H_0(\omega)H_S(\omega)$. Notice that the assumption of a high texture content is common in all focus-related applications since focus measure operators rely on high frequencies in order to detect the focus level. For a discussion about the role of texture in focus-related applications see (Sundaram and Nayar, 1997; Favaro, 2007; Muhammad et al., 2009). The implications of the assumption of a high texture content will be assessed theoretically in section 4.4 and experimentally in section 4.5.4. Fig. 4.8a illustrates the joint effect of diffraction and apodization in the frequency domain. In this figure, the relationship between $H_S(\omega)$ and $H_0(\omega)$ is such that their corresponding cut-off frequencies are close to the system's cut-off frequency, ω_c, in order to obtain the best performance of the acquisition device. As a result, the power spectrum of the band-limited signal, $\tilde{H}(\omega)$, has a *fixed* energy that depends on the diffraction and apodization of the system. In other words, even in the case of perfect focus, the energy of the digital image is bounded by the diffraction and apodization that takes place in the lens-camera system. This effect can be simplified by defining $\tilde{H}$ as:

$$\tilde{H}(\omega) \approx \begin{cases} 1 & \text{if } \omega \leq \omega_e \\ 0 & \text{otherwise} \end{cases} \tag{4.27}$$

where ω_e is the *effective* cut-off frequency of $\tilde{H}$.

It is clear that ω_e is below ω_c (see Fig. 4.8b) and is a system-dependent pa-

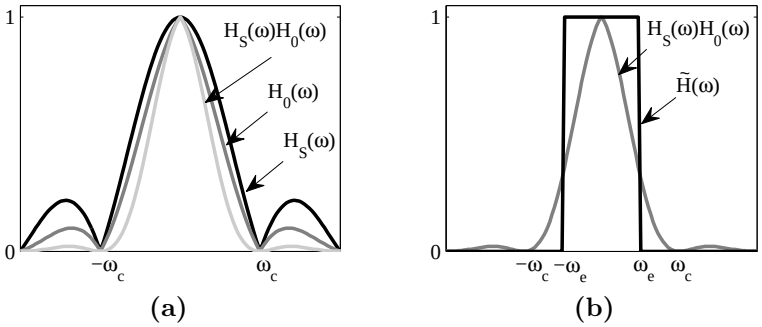


Figure 4.8: Effect of diffraction and apodization in the frequency domain. (a) Diffraction and apodization determine the system's cut-off frequency, ω_c. (b) This effect can be simplified by means of a band-limited approximation of $\tilde{H(\omega)}$.

rameter. Since we are interested in the energy of the captured image as a function of the blur width (i.e., its variation as a function of focus), the approximation in (4.27) suffices for the following analysis provided that $\int |H_0(\omega)H_S(\omega)|^2\, d\omega \approx \int |\tilde{H}(\omega)|^2\, d\omega$. In addition, its simplicity allows modeling the quantitative behavior of defocus while keeping the derived expressions practical and tractable. It is important to remark that, in most off-the-shelf digital cameras, 3rd-order aberrations play an important role that often exceeds that of diffraction. In these cases, this yields a reduced ω_e, but the approximation in (4.27) can still be applied (see appendix C).

Substituting (4.27) and (4.20) in the integral of (4.26), the estimated focus measure, $\tilde{\varphi}$, is defined as:

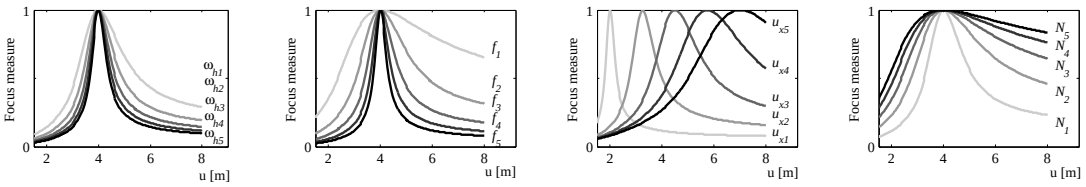
$$\tilde{\varphi} = 2\int_0^{\omega_e} |e^{-\rho^2\omega^2/2}|^2\, d\omega \tag{4.28}$$

The solution to (4.28) is an estimation of the energy of the digital image $I(x, y)$ as a function of the blur parameter ρ. In the ideal case (high-textured noisless image without artifacts or aberrations), the degree of focus, φ, estimated by means of a focus measure operator should vary according to (4.28). In this case, the energy of the transformed image I, after the application of the focus measure operator, is proportional to $\tilde{\varphi}$, so that $\varphi\ \alpha\ \tilde{\varphi}$. Thus, solving (4.25) and normalizing such that $\max\{\tilde{\varphi}\} = 1$, we have:

$$\tilde{\varphi} = \frac{\sqrt{\pi}}{\omega_e\rho}\mathrm{erf}\left(0.5\rho\omega_e\right), \tag{4.29}$$

where $\mathrm{erf}(\cdot)$ is the error function and the blur radius ρ is given by (4.5).

The focus profile model depends on two constants, the effective cut-off frequency ω_e and the camera constant κ; and one parameter, the target position u_x.



(a) $\omega_{h1} < \omega_{hi} < \omega_{h5}$ **(b)** $f_1 < f_i < f_5$ **(c)** $u_{x1} < u_{xi} < u_{x5}$ **(d)** $N_1 < N_i < N_5$

Figure 4.9: Effect of image content and parameters of the lens-camera system on the behavior of the focus profile. (a) Image content, ω_h. (b) Focal length, f. (c) Object position, u_x. (d) f-number, N. Camera parameters: f=35 [mm], N=2.0, s=5e4 [pixel/m].

As stated in the previous section, κ can be found by calibration. In addition, since the effective cut-off frequency is system-dependent, it can be found once by fitting the focus profile of a wide-band image (a high textured image). In this case, the target position, u_x, remains as the only unknown. However, for the sake of generality, ω_e has been left as a parameter in this work (see next section). As a result, if the camera constant is unknown and the effective cut-off frequency is left as a parameter, the proposed focus profile has two unknowns: the target position u_x and the product $\kappa\omega_e$. Interestingly, in contrast to the Gaussian and Laplacian-Cauchy models, these parameters have direct physical meaning: u_x is the target position, κ the camera constant and ω_e is related to the spatial resolution of the system.

4.4 Predicting the behavior of focus

The focus profile model derived in the previous section is based on a wide-band assumption on the source image, since its spectrum is unknown a priori. The general case, without restrictions on the content of the image, is discussed in this section. A commonly used model for an unknown input image corresponds to a band-limited signal (Pratt, 2007):

$$\mathcal{F}\{I_S\} = \frac{A_0}{1 + (\omega/\omega_h)^{2m}}, \tag{4.30}$$

where A_0 is the maximum amplitude of the spectrum of I_S, ω_h is the spatial frequency at half amplitude and m is an integer fall-off factor greater than 1. In order to assess the effect of the image content on the results, Fig. 4.9a plots the focus profile, computed numerically, using (4.30) for different values of ω_h.

From Fig. 4.9a, it is possible to realize that, as ω_h is reduced (i.e., the texture content of the scene decreases), the width of the estimated focus profile increases.

Table 4.1: Effect of image content and parameters of the lens-camera system on focus profile.

i	ω_{hi}	f_i [mm]	u_{xi} [m]	N
1	0.4	20	2.00	1.0
2	0.6	30	3.25	2.0
3	0.8	40	4.50	3.0
4	1.0	50	5.75	4.0
5	1.2	60	7.00	5.0

This is attributable to the fact that the energy of the defocused image is not limited anymore by the resolution of the lens-camera system, but by the cut-off frequency of the imaged scene. As a result, the image content changes the width of the focus peak as a side effect. This is the reason why, in the previous section, we suggest to leave ω_e as a parameter of the model: in order to take into account the effect of the image content in the focus profile.

In practical applications, the reduction of the sharpness of the peak in the focus profile due to the lack of texture implies that the focus peak will be more difficult to be detected. In addition, in agreement with the results in chapter 3, the noise level of the focus profile in real systems is expected to increase since the signal-to-noise ratio (SNR) of the input image is reduced with smooth textures.

As a theoretical contribution, in addition to the effect of the image content, from the proposed focus profile model it is also possible to assess the effect of the parameters of the lens-camera system on the resulting focus profile. For illustration purposes, Fig. 4.9(b)-(c) plots the focus profile for different parameters of the lens-camera system. The parameters used for the simulations are summarized in table 4.1. From this figure, the effect of each parameter can be interpreted as follows:

- *Focal length.* An increasing focal length has two effects on the focus profile: the increased sharpness of the focus peak and the asymmetry of the focus profile. In focus sampling, this implies that, for large focal lengths, the sampling frequency is increased. For SFF applications, this suggests that a large focal length is advisable since it increases the depth resolution of the system. The asymmetry of the focus profile has experimentally been verified in autofocus applications involving short focal length cameras (such as web cameras and cellphone cameras), whereas for large working distances, the focus curve resembles a table top profile (Yousefi et al., 2011; Jeon et al., 2010).

- *Target position.* Both the width and the asymmetry of the focus curve increase as the target position moves away from the camera. This behavior is key in autofocus applications since search strategies (e.g., rule-based search

and hill climbing search) rely on the shape of the focus profile in order to adjust the peak search. In focus sampling this implies that the sampling frequency must be adjusted according to the working distance: the shorter the working distance the higher the required number of focus samples. This is in agreement with the sampling strategy proposed in the previous section since the sampling step Δ_k increases for increasing focus distances.

- *The f-number.* The focus profile flattens out as the f-number increases. This implies that the focus change is more difficult to be detected as the lens diameter is reduced. This agrees with the very well known fact that increasing the lens aperture yields a reduction of the depth-of-field.

As a final remark, the focus measure algorithm used to estimate the focus measure value can also influence the fit between the estimated focus profile and the real one. Depending on their working principle, some algorithms are more or less sensitive to noise. The energy of noise may bias the energy of the pre-filtered image, thus affecting the SNR of the real focus profile. In addition, some focus measure operators apply non-linear transformations to the resulting focus value (such as squaring) in order to sharpen the focus profile. This should be taken into account when applying (4.29) for modeling the focus profile.

4.5 Experiments and discussion

In this section, the focus sampling methodology proposed in section 4.2 is tested. First, the camera calibration procedure is illustrated by calibrating two digital cameras: a Sony RZ50P surveillance camera and a Logitech Orbit AF webcam. In a second set of experiments, the calibrated Sony camera is then used to perform efficient focus sampling and improve the focusing speed in autofocusing tasks. Finally, the focus profile model proposed in section 4.3 is tested and applied to depth estimation through shape-from-focus.

All tests have been performed using two different cameras in order to provide different experimental conditions. On the one hand, the surveillance camera has a high quality and low-distortion optics. In addition, it allows an accurate control of the camera parameters (focal length, aperture, focus position, etc.) and, therefore, theoretical values of the different parameters are available for comparison purposes. On the other hand, the webcam has a higher distortion and optical artifacts. In addition, parameters such as the focal length and pixel size are unknown. Table 4.2 summarizes the characteristics of each camera.

Although the focusing range of both cameras is not upper-bounded, that is, the cameras are able to focus "up to infinity", the maximum focus distance has been limited for practical reasons as shown in table 4.2.

Table 4.2: Technical specifications of the cameras

Camera	Webcam	Surveillance
Focus range	$3 - 24$ cm	$1.5 - 11$ m
Image size	640×480 pix.	640×480 pix.
Focal length	unknown	$9.1e - 2$ m
Pixel size	unknown	$2e5$ pix./m
f-number	unknown	3.8

4.5.1 Focus calibration

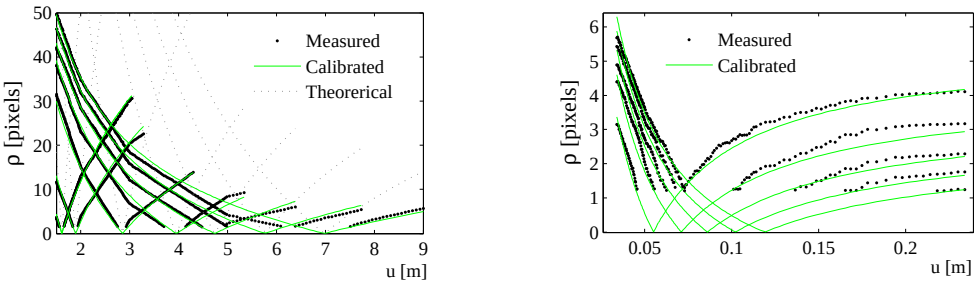
According to the proposed model, the camera constant, κ, does not depend on the position of the calibration target and should be valid for the whole focusing range. In practice, however, changing the focus setting yields a small change in the focal length (section 2.3). Therefore, a more realistic model allows a small variation of κ as a function of the in-focus position:

$$\kappa = Au + B, \tag{4.31}$$

where A and B are constant parameters of the model.

In order to be able to find the parameters in (4.31) and capture their variation as a function of the in-focus position, the calibration process is repeated for at least two different target positions, preferably near the boundaries of the focusing range. For a particular target position, u_{x1}, the corresponding blur width of the imaged target was estimated for a set of images corresponding to different in-focus positions by fitting (4.9) to the intensity profile of the defocused edge step. Finally, a camera constant, κ_1, was estimated by fitting (4.7) to the measured blur widths. The process was repeated in order to compute a camera constant, κ_2, by moving the target to a different position, u_{x2}. The two constants, κ_1 and κ_2, were used to find the model parameters A and B by fitting a straight line to (4.31). For a better accuracy, the process could be repeated for different values of u_x within the focusing range in order to find the model parameters by least squares. In our experiments, using two different target positions yielded acceptable results in terms of accuracy.

The procedure indicated above was carried out for the two cameras described in table 4.2. In general, the above calibration procedure should be easily reproduced for different imaging devices provided that: 1) The digital camera has a motorized (controllable) focus mechanism. In this work, the in-focus distance has been measured in meters but, in general, the model holds for any units (including motor steps). 2) The camera and the scene remain static during the image acquisition stage of the calibration process. The duration of this process may be from several seconds up to minutes depending on the focusing range and the speed of



(a) (b)

Figure 4.10: Measured and predicted blur widths. (a) Calibration parameters for the Sony camera: $A = 7.96$, $B = 82.80$. From left to right, the curves correspond to object positions $u_x =$ 1.75, 2.0, 3.0, 4.0, 5.0, 6.0 and 7.0 m, respectively.(b) Calibration parameters for the Logitech camera: $A = 0.76$, $B = 0.3$. From left to right, the curves correspond to object positions $u_x =$5.5, 7.1, 8.6, 10.2 and 11.9 cm.

the camera. This particular requirement should not represent a major problem. 3) The behavior of the camera can be acceptably described by the first-order Gaussian optics. In general, this applies to the majority of conventional cameras (*e. g.,* photographic cameras, webcams, digital security cameras), whereas third-order aberrations (astigmatism, comma, field curvature, etc) are considered to be either neglected or compensated for so that the measurement of the blur width is not affected.

In order to verify the suitability of the proposed calibration method, the blur width has experimentally been measured for multiple object positions within the calibrated focusing range. The parameter values A and B obtained by calibration have been applied to (4.12) and (4.13) in order to predict the blur widths. Fig. 4.10 illustrates the results for the surveillance camera (Fig. 4.10a) and the webcam (Fig. 4.10b). Taking advantage of the accurate control of the parameters of the Sony camera (its focal length and lens aperture), as well as the manufacturer's information about the imaging sensor (in particular the pixel size) it is possible to obtain a theoretical uncalibrated blur model by means of (4.6). As for the webcam, it is not possible to obtain a theoretical model since both the focal length and the pixel size are unknown.

The uncalibrated model of the surveillance camera (dashed line in Fig. 4.10a) does not correctly describe the observed behavior. This is attributable to the fact that the nominal values of the camera's parameters do not exactly correspond to the real physical values. For instance, although it is possible to know the nominal focal length for a given zoom motor position, the effective focal length of a complex lens system is difficult to be known accurately. In contrast, the calibrated model overcomes the dependency on the camera parameters and correctly conforms to the

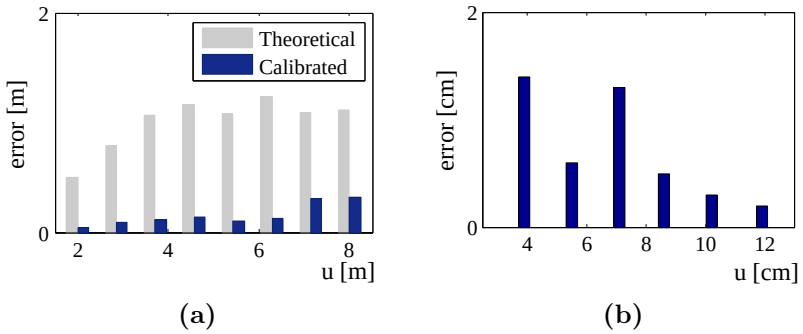


Figure 4.11: Mean absolute estimation error. (a) For Sony RZ50P camera. (b) For Logitech Orbit AF camera (unknown theoretical model).

.

behavior of the real system. The defocus phenomenon, and hence the behavior of the DOF, is an intricate variable that exhibits a non-linear behavior as a function of multiple variables. In most cases, nominal values of the camera parameters are known approximately or may even be unknown for some off-the-shelf cameras, as is the case for the webcam. Interestingly, the proposed model is able to capture the dynamics of the DOF by means of the two calibration parameters A and B.

Fig. 4.11 plots the mean absolute estimation errors as a function of the object position. The estimation error corresponds to the distance between the real in-focus position and the modeled in-focus position for a given blur width (the horizontal distance between the real measured curve and the approximated curves in Fig. 4.10). As shown in Fig. 4.11, the calibrated model predicts the position of the DOF limits with high precision. In particular, the maximum error is not greater than 0.3 m for the Sony camera in a focusing range between 1.5 m and 10 m. The maximum error increases up to 1.2 m (300% higher than the calibrated case) when the blur width is estimated from the nominal values of the camera's parameters without calibration. This experiment shows that, even when the nominal values of the parameters of the lens-camera system are known, there is no guarantee that the theoretical limits of the DOF accurately conform to the real behavior of the system, unless the system is calibrated. For the webcam, the maximum error is not greater than 1.4 cm in a focusing range between 3.9 and 11.9 cm.

The proposed calibration method is simple yet robust to different imaging conditions. On the one hand, the normalization of the intensity profile of the blurred edge compensates for the radiometric response of the camera. On the other hand, it is also well known that, in the presence of defocus, the focus error dominates other third-order optic aberrations. As a result, these aberrations are not expected to deteriorate the outcome of the proposed calibration method in

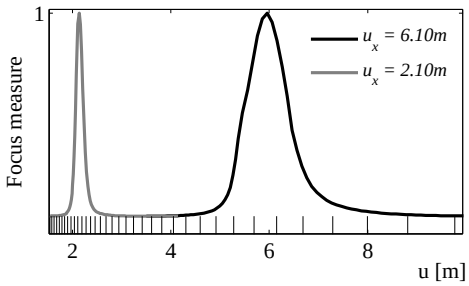


Figure 4.12: Focus sampling in autofocus. For displaying purposes, the focus functions have been normalized between 0 and 1. The vertical lines correspond to the selected focus positions.

normal conditions. In the experiments, optical artifacts and image distortion have been neglected with acceptable results.

4.5.2 Sampling in autofocus

In this section, the focus sampling methodology proposed in section 4.2.3 is exploited in order to boost the speed of autofocusing. The objective of autofocus is to determine the focus setting of a camera in order to bring an observed object into focus. In general, the position of the object with respect to the camera is unknown but, in this work, it will be assumed that the object is located within the focusing range. In order to perform fast autofocusing, it is important to minimize the number of lens positions (focus samples) at which the focus measure operator is applied. A large number of focus samples leads to slow focusing due to both the computation time and the lens movement. Fig 4.12 plots the focus functions corresponding to the same object at two different positions: $u_1 = 2.10$ m and $u_2 = 6.10$ m. In order to find the peak of the shown focus functions, a search strategy should take focus samples at different focus positions. A small sampling step will lead to a slow focusing process, whereas a too large sampling step could override the peak of the functions. As a matter of fact, the selection of the step size is an open problem in autofocus. In addition, as shown in Fig. 4.12, the sharpness of the focus function depends on the position of the object itself. As theoretically predicted in section 4.3, this is due to the change in the focus profile as a function of the target position. Therefore, the sampling frequency should adapt to this behavior accordingly. Without an appropriate model or calibration data, the selection of the sampling step is rather arbitrary or heuristic. In any case, a too small sampling step will lead to a slow process due to oversampling, whereas a too large step could fail to detect objects at certain positions (e.g., the left-most sharp peak in Fig. 4.12).

The horizontal lines in Fig. 4.12 correspond to the positions obtained using

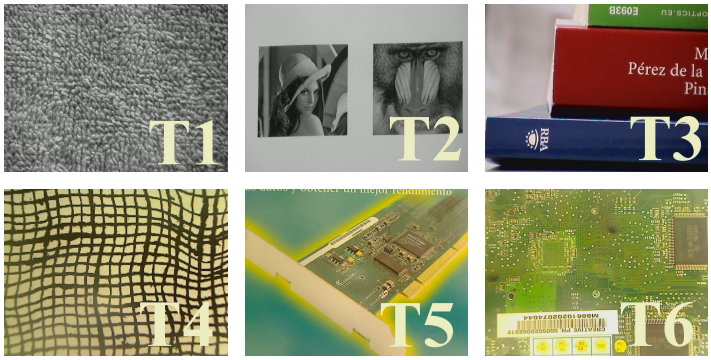


Figure 4.13: Test targets. The focus measure operator has been applied to the central region of the scene (covering 60% of the image). Top row: targets for the surveillance camera. Bottom row: targets for the webcam.

(4.13) and (4.12). It can be observed that the samples effectively adapt to the shape of the focus function and guarantee that there are enough focus samples in order to detect the peak of the focus function, while keeping the number of samples low. In order to compare the proposed approach with Muhammad and Choi (2012), the Nyquist-based sampling frequency is computed as $(u_f - u_n)/12$, where the DOF limits are analytically computed using (2.19) and (2.20).

In the literature, many search strategies have been proposed for autofocus. In this work, both two-step global search and Fibonacci search have been tested[6]. The adaptation of these techniques to the proposed sampling is achieved by simply limiting the search space to the sampling positions computed using the proposed focus sampling methodology. After calibrating the camera, the performance of autofocus has been evaluated using the targets shown in Fig. 4.13. For all the targets, the focus mechanism of the camera was restricted so that there were a maximum of 255 focus steps between the maximum and the minimum focusing distances. The target position u_x and focusing range for each focus sequence are summarized in table 4.3.

The target position, u_x, in table 4.3 is the desired final in-focus position. The autofocus algorithm should yield a final lens position as close to u_x as possible. The difference δ (in meters) between the final lens position and u_x is used as a control variable so that the autofocusing algorithm is considered to have failed if this difference yields an appreciable blur. The performance measure corresponds to the number of iterations (steps) of the autofocus algorithm required for finding the focus peak, S. As for the Fibonacci search, the starting lens position for each target is randomly selected. The results corresponding to the average of 500 trials

[6]Hill-climbing and rule-based search were not included due to their dependence on different heuristic parameters. These parameters depend on the sampling step and, hence, would not allow an objective comparison

Table 4.3: Focusing range and target position of test targets for autofocus. u_x is used as a control variable in order to compute the autofocusing error, δ.

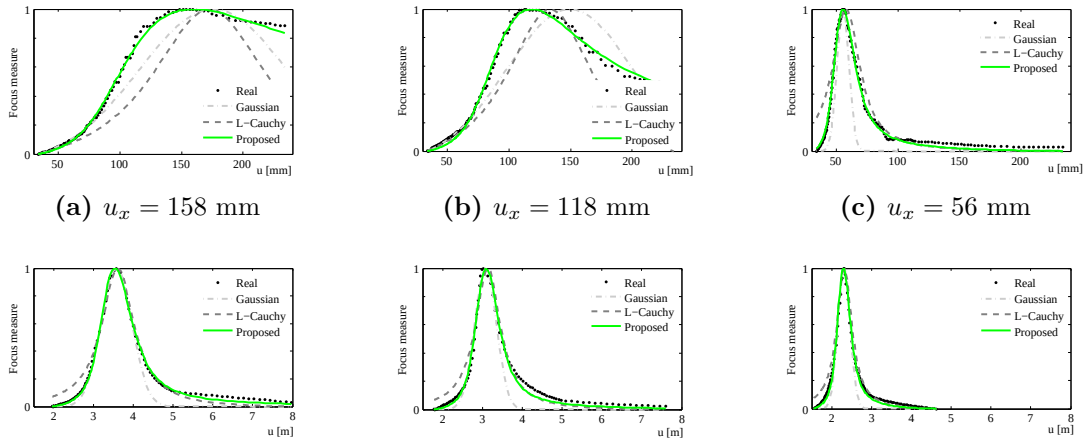
Target	$u_{\min}$ [m]	$u_{\max}$ [m]	u_x [m]
T1	1.57	2.14	1.81
T2	3.19	8.90	4.66
T3	2.94	10.00	5.16
T4	0.03	0.234	0.06
T5	0.03	0.234	0.07
T6	0.03	0.234	0.08

Table 4.4: Performance of autofocus (S) with different sampling strategies: global search and Fibonacci search (in parenthesis).

	Steps (S)			
Target	Full	Nyquist	Theoretical	Calibrated
T1	70 (12)	64 (11)	28 (9)	11 (6)
T2	70 (12)	93 (12)	32 (10)	14 (7)
T3	70 (12)	91 (12)	37 (10)	14 (7)
T4	130 (12)	-	-	7 (5)
T5	130 (12)	-	-	7 (5)
T6	130 (12)	-	-	7 (5)

for each test target are summarized in table 4.4. Four sampling approaches have been tested: Using the whole image sequence (*Full*), Using the Nyquist-based approach proposed by Muhammad and Choi (2012) (*Nyquist*), using the proposed sampling with the nominal values of the lens-camera system without calibration (*Theoretical*) and after calibration with the proposed method (*Calibrated*).

In all the experiments reported in table 4.4, the error δ for all targets, sampling methods and search strategies was always below 0.05m for T1-T3 and below 0.002 m for T4-T6. In the tested focusing ranges, this error yields negligible differences with respect to the amount of focus. It is clear from this table that the proposed calibrated focus sampling significantly reduces the necessary number of steps required to reach the maximum focus. This is explained by the fact that the proposed method covers the focusing range with the minimum number of focus samples. In addition, the focus samples are distributed with a sampling frequency that changes according to the width of the focus function. Notice that the theoretical sampling without calibration still yields an oversampled focusing process. From Fig. 4.10a it is clear that the theoretical uncalibrated model underestimates the location of the near and far limits of the DOF (i.e., the real blur width is far below the estimated values). As a result, the focusing algorithm captures more



(a) $u_x = 158$ mm **(b)** $u_x = 118$ mm **(c)** $u_x = 56$ mm

(d) $u_x = 3.6$ m **(e)** $u_x = 3.0$ m **(f)** $u_x = 2.3$ m

Figure 4.14: Focus profiles of targets at different positions approximated by different focus profile models. (a)-(c) Profiles for webcam. (d)-(e) for the surveillance camera.

focus samples than necessary. However, the opposite case can also be expected in real applications: The theoretical model, without calibration, can overestimate the blur width. In this case, the focusing algorithm will capture fewer images than necessary, which can lead to a failure when finding the focus peak.

4.5.3 The focus profile

In this section, the focus profile model derived in section 4.3 is compared with both the Gaussian profile (4.16) and the Laplacian-Cauchy profile (4.17) for fitting the real focus profile. The root-mean-squared-error (RMSE) and the correlation (C) between the fitted model and the measured focus profile have been used as *quality measures*. The comparative tests have been performed as follows.

A target object at a fixed position u_x is placed in front of a camera and a focus sequence is captured by changing the focus of the camera. The real focus profile is then measured for 100 different random image coordinates using a region of interest of 64×64 pixels for the computation of the focus measure. The compared focus profile models are fitted and the quality measures are computed and averaged over the 100 experiments. This process is repeated for six different object positions by using the different acquisition devices: the Sony SNC-RZ50P surveillance camera and the Logitech Orbit AF webcam. The obtained results are summarized in Table 4.5.

According to Table 4.5, the proposed focus profile outperforms the alternative models for both acquisition devices, different target positions and quality measures. The performance of the compared focus profile models is illustrated in Fig.

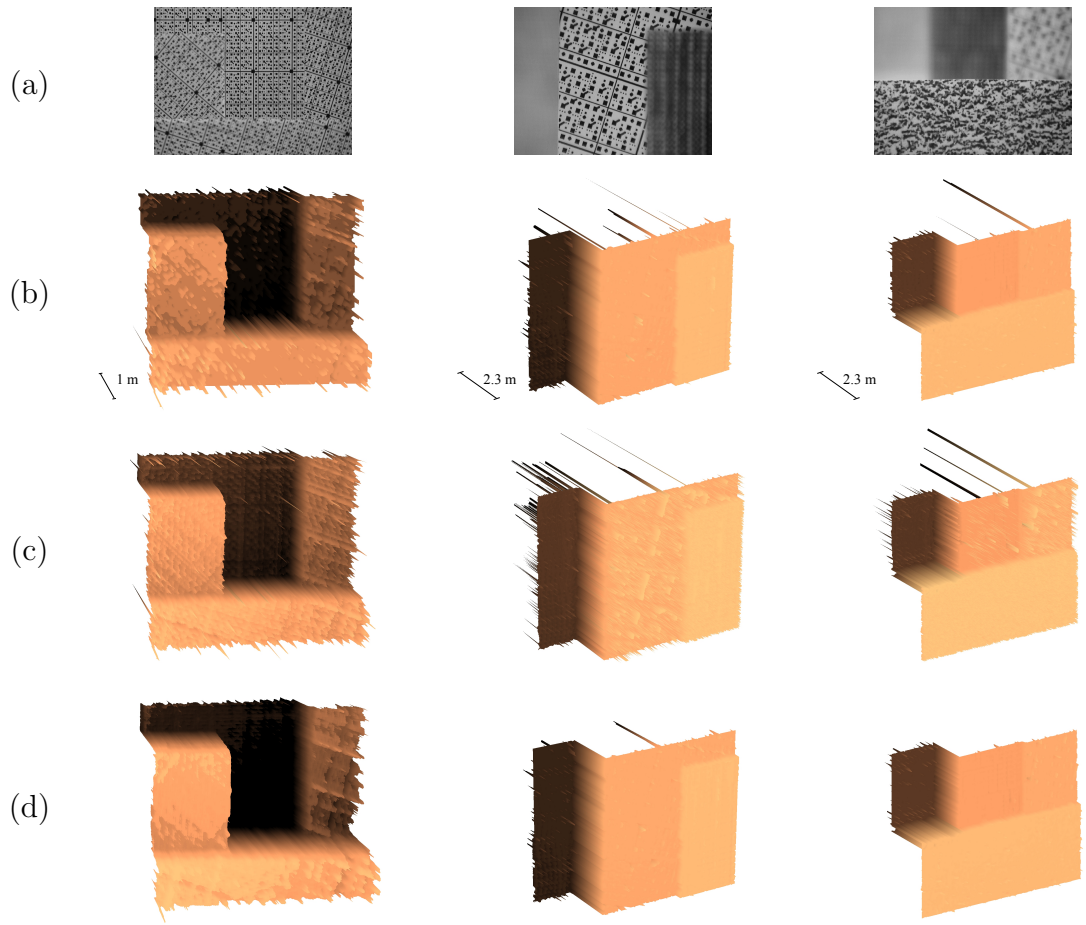


Figure 4.15: Performance of shape-from-focus using different focus profile models. (a) One frame of the focus sequence. (b) Gaussian model. (c) Laplacian-Cauchy model. (c) Proposed model.

Table 4.5: Mean quality measures of the different focus profile models as a function of the target position.

(a) Webcam

	Gaussian		Laplacian-Cauchy		Proposed	
u_x [m]	RMSE	C	RMSE	C	RMSE	C
0.158	0.083	0.984	0.155	0.949	0.016	0.999
0.138	0.116	0.963	0.174	0.927	0.020	0.999
0.118	0.119	0.942	0.157	0.933	0.024	0.998
0.080	0.297	0.588	0.151	0.896	0.032	0.997
0.056	0.205	0.896	0.135	0.941	0.024	0.997
0.042	0.139	0.944	0.117	0.952	0.022	0.998

(b) Surveillance camera

	Gaussian		Laplacian-Cauchy		Proposed	
u_x [m]	RMSE	C	RMSE	C	RMSE	C
4.7	0.139	0.934	0.119	0.942	0.061	0.985
4.2	0.082	0.979	0.072	0.978	0.048	0.991
3.6	0.068	0.989	0.062	0.981	0.020	0.998
3.0	0.094	0.964	0.074	0.967	0.033	0.994
2.6	0.087	0.979	0.085	0.976	0.047	0.994
2.3	0.076	0.982	0.069	0.985	0.027	0.997

4.14 for different target positions and acquisition devices. From this figure, it is possible to realize that the proposed model adjusts more accurately to the real focus profile than the other two models. This is particularly evident with the profiles corresponding to the webcam (Fig. 4.14a-c). For the surveillance camera, this is less evident since its large focal length yields more symmetric and sharper focus profiles.

In a practical application, the benefits of the proposed focus profile model can be verified in depth estimation through shape-from-focus. In this scope, the proposed focus profile model is used to interpolate the focus function in the peak detection stage of the traditional shape-from-focus framework. In order to simplify the construction and measurement of the ground truth, the captured scenes mostly consisted of planar objects placed at different positions from the camera. For each scene, a focus sequence was acquired using the surveillance camera. The depth-map was then estimated by using the classical SFF framework described in section 2.2.3.

Five sequences have been used in the experiments and the depth-maps generated with different focus profile models have been compared. In order to measure

Table 4.6: Performance measures of SFF with different focus profile models: Gaussian profile GP (Nayar and Nakagawa, 1994), Laplacian-Cauchy profile LC (Muhammad and Choi, 2012; Tsai and Chen, 2012) and proposed focus profile model FP.

Method	RMSE	AE (%)	SNR (dB)	C
GP	0.108	1.52	35.9	0.972
LC	0.178	2.51	33.4	0.932
FP	0.095	1.41	36.6	0.975

the quality of the obtained depth-maps, the RMSE, the correlation (C), the signal-to-noise ratio (SNR), and the mean absolute error (AE) have been computed between the reconstructed depth-map, $z(x, y)$, and the ground truth, $z_G(x, y)$. The mean absolute error is defined as:

$$AE(\%) = 100 \sum_{\forall(x,y)} \frac{|z(x, y) - z_G(x, y)|}{z_{GT}(x, y)} \tag{4.32}$$

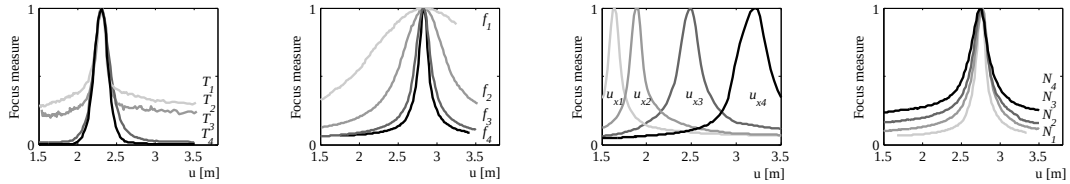
Fig. 4.15 shows one frame of the focus sequence and the obtained depth-maps for three of the sequences used in the experiments. The mean results are summarized in Table 4.6. As shown in this table, the best performance of SFF is obtained when the proposed focus profile model is applied.

4.5.4 Focus measure behavior

In section 4.4, the proposed focus profile model was used to theoretically predict the effect of changing the parameters of the lens-camera system on the behavior of the real focus profile. In this section, the claims in section 4.4 are experimentally verified. For the case of the f-number, N, the lens focal length, f, and the target position, u_x, this is accomplished by simply changing the corresponding parameter of the camera and measuring the real focus profile for each configuration. For the image content, since it is not possible to have a strict control of the frequency content of the captured scene, targets with different textures have been used, with four of them having been selected for illustration purposes. The real parameters of the lens-camera system are summarized in Table 4.7. Figure 4.16 plots the focus profiles obtained for different conditions. The textures used to generate Fig. 4.16a are shown in Fig. 4.17.

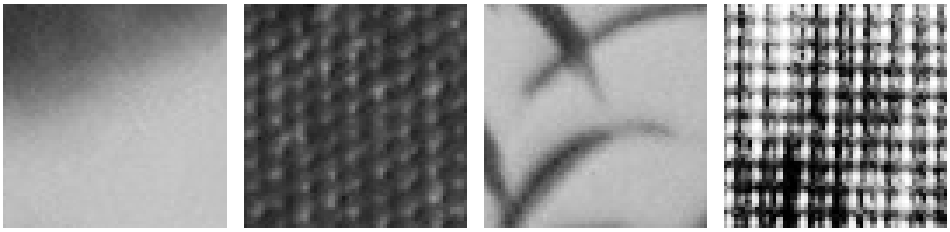
By comparing Fig. 4.9 and Fig. 4.16, the following effects can be verified:

- *Focal length.* Both the sharpness and the symmetry of the focus profile increase with increasing focal lengths.



(a) see table 4.7 **(b)** $f_1 < f_i < f_4$ **(c)** $u_{x1} < u_{xi} < u_{x4}$ **(d)** $N_1 < N_i < N_4$

Figure 4.16: Effect of image content and parameters of the lens-camera system on the focus profile. (a) Image content. (b) Focal length, f. (c) Target position, u_x. (d) f-number, N. Camera parameters: f=91 [mm], N=1.6, s=6e4 [pixel/m].



(a) T_1 **(b)** T_2 **(c)** T_3 **(d)** T_4

Figure 4.17: Textures used to assess the effect of the image content on the focus profile.

Table 4.7: Effect of image content and parameters of the lens-camera system on the focus profile.

i	f_i [mm]	u_{xi} [m]	N
1	91.0	1.64	1.6
2	65.1	1.88	4.0
3	36.4	2.49	6.8
4	22.8	3.20	9.6

- *Target position.* The width of the focus profile decreases as the target moves away from the camera. The focus profile tends to the table-top profile as the working distance increases.

- *f-number.* Large lens apertures reduce the sharpness of the focus peak.

With regard to the effect of the image content on the obtained focus profile, it can also be verified in Fig. 4.16a that a reduction in the frequency content of the captured image yields a reduced focus peak. However, there are two additional effects that are not accurately predicted by the proposed focus profile model when the texture content of the target is reduced:

1. The smoothness of the curve is deteriorated.

2. The DC component of the curve increases.

Both effects can be attributed to a single cause: the image noise. On the one hand, as demonstrated by Subbarao and Tian (1998), the variance of the real focus profile is a function of the image noise. Thus, if the imaged target has a low texture content, the SNR decreases. As a result, the contribution of noise to the variance of the normalized focus profile is more evident. On the other hand, the addition of noise to the captured image takes place after the low-pass filtering effects of the system's optics (i.e., noise is mostly electronic). As a result, noise is a wide-band component that increases the energy of the captured image and is present regardless the degree of focus, manifesting as a DC component on the focus profile. Therefore, the contribution of noise to the normalized focus profile is more evident when a low-energy (weakly-textured) target is imaged. The fact that the proposed focus profile model fails to predict this behavior is not unexpected since, in its derivation, the effect of noise has been neglected. Interestingly enough, this is in agreement with the experimental results and the analysis about the effect of image noise in the estimation of the focus measure presented in chapter 3.

Another factor that indirectly affects the focus profile is the focus measure operator itself. In practice, since different focus measure operators transform the input image and modify their frequency content in a different manner, the effective cut-off frequency ω_e may differ for the same imaging conditions. In this work, preliminary experiments have been performed using focus measure operators based on different principles, such as the image Laplacian, image gradient, wavelet transform and image statistics. Similar qualitative results are obtained by varying the focus measure operator depending on its robustness to noise, with only minor differences in both the estimated $\kappa\omega_e$ parameter and the smoothness of the focus profile curve.

4.6 Summary

The contributions in this chapter can be summarized as follows.

- A robust and efficient method for the calibration of the focus of a camera that can be exploited for efficient focus sampling.

- A theoretical focus profile model.

The calibration method presented in section 4.2.2 is aimed at estimating the parameters that describe the behavior of the camera's focus as a function of the lens configuration. The main advantage of the proposed calibration method is twofold: first, it does not rely on any information about the physical parameters of the camera-lens system, namely the lens focal length, the effective pixel size or the numerical aperture. Therefore, it can be applied in the general case when some/all of the parameters of the lens-camera system are unknown. Secondly, it is robust and readily carried out by performing a focus sweep on an edge target. Once the calibration is performed, the focusing process can be regarded as a variable-resolution camera whose resolving power has a closed form as a function of the calibration parameters.

The proposed calibration method allows an accurate prediction of both the near and far limits of the depth-of-field. This knowledge can then be exploited for performing efficient focus sampling (section 4.2.3). The proposed sampling strategy was applied in order to speed up the autofocus process for different cameras. The obtained results summarized in table 4.4 show that the proposed method significantly reduces the number of frames required to reach the in-focus position.

The theoretical focus profile model presented in section 4.3 aims at describing the focus measure value as a function of the focus position. The proposed model has been derived by considering some simplifications of the image formation process and by assuming that the focus measure is proportional to the energy of the ideal blurred image. A fundamental aspect of the proposed focus profile model is that it does not require the accurate estimation of the PSF of the defocused system. Instead, the problem is formulated as finding the energy of the system as a function of defocus. In our experiments, the advantages of a simple model with an analytical expression for the focus profile compensate for the limitations of the model itself. The proposed model has been compared to previous alternatives: the Gaussian profile (GP) and the Laplacian-Cauchy (LC) profile, in order to fit the real focus profile for two different acquisition devices. The results clearly show that the proposed focus profile model outperforms the other two models both for describing the real focus function in conventional cameras and for depth estimation through shape-from-focus.

In addition to its good performance, an advantage of the proposed focus profile model is that its parameters have a real physical interpreation. In its most general form, the model depends on two parameters: the target position, u_x, which is a function of the scene geometry and the product $\kappa\omega_e$, which is a function of the parameters of the lens-camera system. On the one hand, κ groups in a single value the joint effect of the lens focal length, f, the aperture number, N, and the pixel size, s. On the other hand, the effective cut-off frequency, ω_e, accounts for the resolving capability of the lens-camera system at the diffraction limit. The physical meaning of the model's parameters was exploited in section 4.4 in order to perform a qualitative theoretical assessment of the effects of the different parameters of the acquisition device on the resulting focus profile. This analysis was experimentally validated in section 4.5.4.

Having a theoretical model that describes the behavior of the focus function in conventional cameras can be further exploited for assessing the confidence of the focus estimation in real imaging conditions (see the next chapter). In addition, a closed-form expression for the focus profile paves the way for new self-adaptive approaches for efficient image acquisition and autofocus in limited depth-of-field systems.

CHAPTER 5

Confidence of the focus estimation

An important problem that affects the estimation of the focus level and, therefore, the performance of different focus-based applications is the *image content problem*. Most research in focus-based applications has been devoted to improving the robustness of focus estimation by means of new focus measure operators or by improving the obtained results through different post-processing techniques. Alternatively, this chapter proposes a reliability measure (R-measure) aimed at assessing the confidence of the estimated focus measure value. This valuable information is exploited for removing corrupted data in the depth-maps generated through shape-from-focus and for generating noise-robust all-in-focus images through focus stacking. Unlike previous approaches, the proposed R-measure integrates efficiently into the focus estimation task for addressing the image content problem, without any previous knowledge about the content of the imaged scene or the associated depth-map.

This chapter is organized as follows. Section 5.1 describes in more detail the image content problem and reviews the approaches proposed in the literature in order to overcome it. The proposed R-measure is described in section 5.2. In a practical application, the R-measure is used for depth-map carving in shape-from-focus in section 5.3, and for all-in-focus image generation through focus stacking in section 5.4. The proposed approach is experimentally evaluated in section 5.5. The obtained results are summarized in section 5.6

5.1 Introduction

In the literature, most focus-based applications rely on focus measure operators for the detection of high-contrast regions in an image stack. Notwithstanding, as stated in previous chapters, a difficulty in the measurement of the focus level is the lack of strong texture in the imaged scene, namely the *image content problem*. The trivial example is a completely textureless plane in front of the camera. In this case, with varying focus, the only appreciable variations on the imaged scene are mostly due to side effects of image electronic noise and possibly slight illumination variations due to the movement of the lenses. This unability of focus measure operators to detect focus variations in this case was theoretically demonstrated by Sundaram and Nayar (1997). Subsequently, Favaro et al. (2003) demonstrated that images that satisfy the Laplace equation $\Delta I = 0$ are invariant to any circularly symmetric PSF and can not induce detectable changes in the focus measure.

An undesired side effect related to the image content problem is that focus measure operators can erroneously assign relatively high focus measure values to regions with low signal-to-noise ratio (SNR), due to the high frequency components of noise. For the particular case of focus stacking, this yields all-in-focus images with increased noise. In turn, in focus-based depth estimation (e.g., in shape-from-focus and shape-from-defocus), the incorrect estimation of the focus measure yields inaccurate depth-map generation.

The image content problem in depth estimation is illustrated in Fig. 5.1. This figure shows a synthetic scene that consists of a conical surface with a texture mapped on it. Fig. 5.1b shows the corresponding depth-map obtained through shape-from-focus. It is clear that some regions of the obtained reconstruction are inaccurate and highly corrupted by noise. In this case, traditional smoothing techniques such as median filtering (Nayar and Nakagawa, 1994), bilateral filtering (Tomasi and Manduchi, 1998) or non-local means (Buades et al., 2005a,b) are of limited application since large areas of the recovered scene are unreliable (see Fig. 5.1c-5.1e). Depending on the application, there are more sophisticated approaches that can be exploited in order to compensate for unsatisfactory results due to deficiencies of the focus measure operators when detecting focus. In the example of Fig. 5.1, the obtained depth-map can be improved by post-processing it with different filtering, denoising, regularization or smoothing techniques (Mahmoudi and Sapiro, 2012). Alternatively, the technique proposed in this chapter is aimed at measuring the reliability of the focus measure estimation according to the behavior of the focus function. This reliability measure (R-measure) is used to indentify the regions (pixels) where the depth estimation is likely to be incorrect in order to discard them (depth-map carving). As for focus stacking, the R-measure is exploited for performing a selective image fusion in order to generate noise-robust

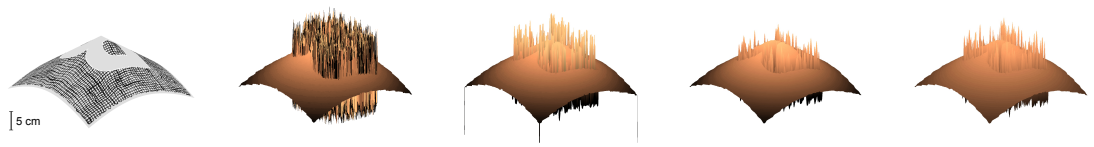


Figure 5.1: The image content problem in shape-from-focus. (a) Synthetic scene. (b) Depth-map obtained through SFF. (c) Smoothing with median filter: Nayar (1989); Nayar and Nakagawa (1994). (d) Smoothing with bilateral filter: Tomasi and Manduchi (1998). (e) Smoothing with non-local means: Buades et al. (2005a,b).

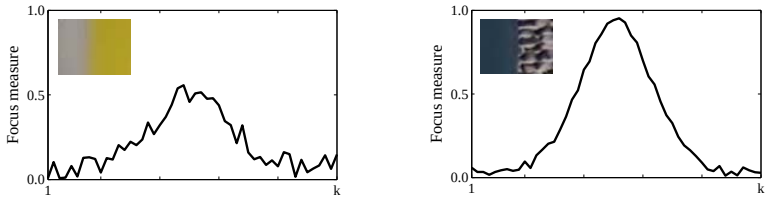
all-in-focus images.

Previous work

In practice, the influence of the image content on the focus measurement has been studied mostly for the estimation of depth-maps through shape-from-focus. In that scope, Shoji et al. (2006) used color segmentation and bilateral filtering to improve both the accuracy of the focus measurement and the final estimation of depth. The main drawback of this approach is that the region merging stage is performed by taking into account color and not texture, which is the key factor for the focus measure. This may lead to erroneous results for different objects and surfaces with the same color. The approach proposed by (Shoji et al., 2006) does not provide information about the reliability of the estimated focus measure of the image pixels.

In (Muhammad et al., 2009), a depth-map is initially obtained using classical shape-from-focus. Then, parts of the scene with high depth variations are discarded in considering that they are due to an inaccurate computation of the focus measure. The discarded regions are then recovered by interpolation. The disadvantage of the latter approach is that it is only applicable to scenes where non-reliable regions are small and can be interpolated from highly-textured ones. Alternatively, Gaganov and Ignatenko (2009) apply Markov random fields in order to smooth the obtained depth-map in low-reliability areas. This approach also assumes that the depth information of highly-textured areas can be used to infer and constrain the depth estimates where shape-from-focus fails, but it does not provide information about the location of low- and high-reliability areas. More recently, Muhammad and Choi (2011) have proposed to carve the depth-map by applying a Canny edge detector to the all-in-focus image of the scene.

Within the scope of this dissertation, the main advantage of the proposed R-measure is that it exploits the information of the focus measure values of the focus



(a) (b)

Figure 5.2: Behavior of focus measure according to the image content. (a) Low-textured scene. (b) High-textured scene.

stack. Since the computation of the focus stack is a fundamental part of several focus-related applications, the R-measure can be efficiently integrated into these frameworks. Unlike previous approaches, the proposed methodology computes the reliability measure without the need for either computing the all-in-focus image or post-processing the generated depth-map. The experiments on both synthetical and real data sets show the advantages in performance and efficiency of the proposed approach with respect to state-of-the art alternatives.

5.2 Proposed reliability measure

According to previous chapters, there are different factors that affect the focus level estimated by means of focus measure operators, such as the image noise, the size of the support window and the parameters of the acquisition device. In addition, the texture content of the imaged scene can also affect the estimation of the measure. As a result, predicting whether the estimated focus value is reliable or not is a rather involved task. In the research community, most effort has been devoted in proposing new robust focus measure operators or devising new post-processing techniques for improving the obtained results. However, by taking advantage of the results on focus profile modeling of the previous chapter, it is indeed possible to determine if the measured focus profile *conforms* to the expected one. This concept is illustrated in Fig. 5.2, where the focus functions corresponding to different image regions are plotted. From those curves, it can be appreciated that the focus functions corresponding to regions with different texture patterns exhibit different behaviors. On the one hand, the image features in Fig. 5.2a are "weak" when compared to the noise present in the sequence. On the other hand, the image texture in Fig. 5.2b stands out over the existing noise, leading to a clear response with a maximum at the position of highest focus.

Let the focus function for a pixel at coordinates (x, y) be a signal that varies

according to both the degree of focus of this pixel plus an *error signal*:

$$\varphi_{x,y}(u) = \tilde{\varphi}_{x,y}(u) + E_{x,y}(u), \tag{5.1}$$

where $\varphi_{x,y}$ is the measured focus function for that pixel, $\tilde{\varphi}_{x,y}$ the associated ideal focus profile and $E_{x,y}$ an error signal that represents the departure of the focus function from the ideal behavior. $E_{x,y}$ may be explained by image noise, lack of texture, limitations of the focus measure operator or departure of the focus function from its ideal behavior. The ideal theoretical focus profile, $\tilde{\varphi}$, can be computed as typically done in SFF, by fitting an analytic focus profile model to the measured focus function.

Depending on the particular application, the analytical focus profile used in (5.1) can be any of the analytical models presented in the previous section: a Gaussian function, a Laplacian-Cauchy function or the focus profile model in (4.29). Therefore, this scheme can be readily adapted to non-conventional cameras, such as microscopes. Despite the chosen model, it is clear that the reliability of the focus measure estimation depends on how well the ideal focus profile, $\tilde{\varphi}$, approximates the measured focus function, φ. More specifically, the variance of the measured focus function can be attributed to the error signal. Bearing this in mind, a reliability of the focus measure corresponding to a pixel at coordinates (x, y) is defined as:

$$R(x,y)^{-1} = \frac{1}{K f_{\max}} \sum_{\forall k} |\varphi_{x,y}(u_k) - \tilde{\varphi}_{x,y}(u_k)| \tag{5.2}$$

$$R(x,y)^{-1} = \frac{1}{K f_{\max}} \sum_{\forall k} |E_{x,y}(u_k)|, \tag{5.3}$$

where $f_{\max} = \max\{\varphi_{x,y}\}$ is a normalization factor.

The R-measure in (5.3) has high values ($R \to \infty$) for weak error signals and low values ($R \to 0$) for strong error signals. For convenience, let:

$$e_{x,y} = \frac{1}{K} \sum_{\forall k} |E_{x,y}(u_k)|, \tag{5.4}$$

denote the absolute average of the error signal[1].

In 5.3, the R-measure has been expressed analogously to the signal-to-noise (S/N) ratio. One valuable benefit of S/N analysis is the possibility to state the confidence limits (in units of standard deviations) that the focus measure signal

[1] Alternatively, the components of the error signal can be added in quadrature yielding similar results, since $\sum |E_{x,y}(u_k)| \approx \sqrt{\sum} |E_{x,y}(u_k)|^2$

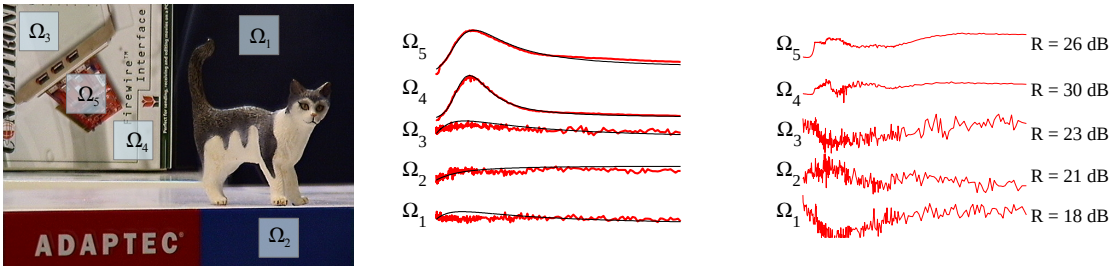
stands out from the noise (error signal). In this sense, the error signal, $e_{x,y}$, can be interpreted as an approximation of the *relative error*, that is, the fractional component of the signal that is noise. As a result, the ratio $f_{\max}$ over $e_{x,y}$ describes the confidence level at which a focus function of a certain strength can be distinguished from noise (Murphy, 2001, chapter 15). The reliability measure can then be expressed analogously to the PSNR by taking the logarithm in (5.3), yielding:

$$R(x,y) = 20\log\left(\frac{f_{max}}{e_{x,y}}\right), \tag{5.5}$$

As previously stated in chapter 3, some researchers have exploited different features of the focus function, such as the energy of the noise signal, the number of false maxima, and the sharpness and height of the focus peak, for assessing the performance of focus measure operators. In preliminary experiments, the features of the focus function were also studied as reliability measures (either individually or combined by means of neural networks). The best results were obtained with the R-measure in (5.5). This is not surprising since, in chapter 3 it was experimentally proved that the features of the focus function are only weak predictors of the performance of the focus measure operator. For the sake of brevity, only the R-measure is described here.

Figure 5.3 corresponds to a simple experiment that illustrates the working principle of the proposed R-measure. Fig. 5.3a is the all-in-focus image corresponding to a focus sequence of 255 frames with five regions of interest (ROI) being selected $\{\Omega_i | i = 1, \cdots, 5\}$. The ROIs have been chosen in order to include both highly-textured regions and low-textured regions of the scene. From Fig. 5.3b, it can be realized that, as the *strength* of the image signal is reduced, the focus measure estimation is less accurate and, therefore, the fit between the measured focus function, φ, and the corresponding theoretical focus profile model, $\tilde{\varphi}$, deteriorates. As a result, the energy of the error signal (Fig. 5.3c) increases yielding a reduced R-measure value.

Although the design and development of new focus measure operators capable of reliably estimating the focus level under different imaging conditions is an important concern, we are also interested in being able to determine if the estimated focus measure is reliable or not. In this sense, the focus profile model developed in the previous chapter provides a reference benchmark that allows one to determine if the imaging process is close to or departs from the ideal behavior. Beyond being a simple indicator of how ideal the measured focus function is, the R-measure can be successfully exploited in depth-map carving in shape-from-focus and for the generation of all-in-focus images through focus stacking. These applications are discussed in more detail below.



(a) (b) (c)

Figure 5.3: Computation of the reliability measure. (a) All-in-focus image of a focus sequence with five regions of interest being selected. (b) Measure focus function (red) and the corresponding theoretical focus profile model (black). (c) Error signal and reliability measure.

5.3 Efficient depth-map carving

It is straightforward to apply the proposed R-measure for determining the regions where the focus estimation is unreliable in order to identify the pixels where depth estimation techniques, such as shape-from-focus, will not perform correctly. In particular, the values of the R-measure for all pixels can be interpreted as a gray-scale image in which each gray level is associated with the reliability of the depth estimation for the corresponding pixel: the depth of those pixels with a high R-measure is more likely to be estimated correctly. As a result, the depth-map can be carved by removing the pixels whose depth estimation are associated with a low R-measure. In particular, all pixels whose reliability is below a predefined threshold, $R_{\min}$, should be discarded.

The reliability threshold must be found experimentally after a training process. For a given training set, the depth-map carving can be thought of as a two-class classification task where the classes correspond to those pixels that should be discarded from the depth-map and those that should be kept. Thus, the threshold value $R_{\min}$ is selected so that the highest classification rate in the training set is obtained in order to maximize the classification accuracy. As in typical classification tasks, accuracy corresponds to the percentage of correctly classified pixels with respect to the total number of pixels in each image. Specifically, a carving mask, M, is generated as:

$$M(x,y) = \begin{cases} \text{true} & \text{if } R(x,y) < R_{\min} \\ \text{false} & \text{otherwise} \end{cases} \tag{5.6}$$

Fig. 5.4 illustrates the proposed depth-map carving methodology for the scene previously shown in Fig. 5.1. Fig. 5.4a corresponds to the reliability map associated with the corresponding focus sequence. The light colors correpond to pixels

(a) (b) (c)

Figure 5.4: Depth-map carving through reliability measure. (a) Reliability map. The lighter the color the higher the R-measure. $R(x, y) \in [0, 39]$ dB. (b) Depth-map carving. White pixels correspond to an R-measure below the threshold. (c) Carved depth-map.

with high reliability measures. The depth-map carving is performed by selecting the pixels whose reliability is below a given threshold (the white pixels in Fig. 5.4b). The resulting depth-map after carving is shown in Fig. 5.4c. The application of the proposed R-measure for depth-map carving is experimentally validated in section 5.5.1.

5.4 Improved focus stacking

In addition to depth-map carving, the R-measure can be exploited for adaptive image fusion in focus stacking. In this scope, a straightforward solution to the image fusion problem is to compute the intensity of a pixel in the all-in-focus image ψ at coordinates (x, y) as a weighted average of the pixels in the original sequence (Sugimoto and Ichioka, 1985). The weights will be proportional to the the focus measure value:

$$\psi(x, y) = \frac{1}{F_n} \sum_{\forall k} F_k(x, y) I_k(x, y), \tag{5.7}$$

where F_n is the is a normalization factor such that $\frac{1}{F_n} \sum_{\forall k} F_k(x, y) = 1$.

Notwithstanding, this approach has two drawbacks: first, a linear combination of all frames yields a low-contrast, all-in-focus image. Second, the sensitivity of all focus measure operators to high-frequency components in the images will yield a low-quality all-in-focus image in the presence of noise. The latter problem can be explained by the fact that the *energy maximization scheme* (section 2.2.4) cannnot distinguish when the high energy associated with a pixel is due to the signal itself or to image noise. As a result, the fusion process sharpens the pixels that correspond

to both image features and noise artifacts, without distinction. This problem is common to most all-in-focus algorithms previously reviewed in section 2.2.4.

In this section, a selective focus stacking algorithm is proposed. The algorithm works in two stages. In the first stage, a sharpening parameter is computed based on the R-measure. The aim of the sharpening parameter is to assign a high value to those pixels that exhibit a high focus measure value that is likely to be attributed to a strong visual pattern in the scene. In contrast, those pixels with either a high focus measure value associated with high noise or with low focus measure values attributed to low texture content in the scene are assigned a low sharpening index. In the second stage, the all-in-focus image is generated from the source image stack according to the sharpening parameter. These stages are described below.

5.4.1 Sharpening index

In order to perform a selective image fusion for generating a low-noise all-in-focus (AIF) image, the focus measure must be complemented with a selection scheme that allows the system to determine if the focus measurement is reliable or not, by means of the previously proposed R-measure. For this purpose, the *sharpening index* ϕ is defined as a function of the R-measure as:

$$\phi(x,y) = \alpha\left(\frac{1}{2} + \frac{1}{2}\tanh(\alpha^{-1}(R - R_{th}))\right), \tag{5.8}$$

where the reliability threshold R_{th} and the scaling constant α are parameters of the proposed focus stacking algorithm that must be determined experimentally.

The aim of the sharpening index in (5.8) is to provide a bounded response as a function of the R-measure. Notice that the R-measure is unbounded, that is, it can have any value in $[-\infty, +\infty]$. This can yield computational unstability when computing the AIF image as a function of ϕ. Alternatively, the sharpening index in (5.8) provides a linear response as a function of the R-measure for values near the reliability threshold R_{th} and saturates between $[0, \alpha]$ for values away from the reliability threshold (Fig. 5.5). Thus, as illustrated in Fig. 5.5, the sharpening index assigns high values ($\phi \to \alpha$) to pixels where the measured focus function conforms to the theoretical focus profile (i.e., $\varphi \approx \tilde{\varphi}$). In contrast, pixels where the measured focus function departs from the theoretical focus profile ($\varphi \neq \tilde{\varphi}$) are assigned a low value ($\phi \to 0$). Both the selection of the reliability threshold and the scaling constant will be discussed in more detail in section 5.5.2.

5.4.2 Image fusion

In the final step of the proposed focus stacking algorithm, an image fusion process is performed according to the *activity* of the image pixels, the latter estimated by

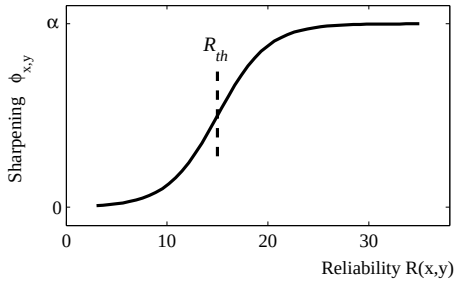


Figure 5.5: Sharpening index as a function of reliability. The sharpening index is a linear function of the reliability around R_{th} and saturates between 0 and α for low and high reliability values, respectively.

means of both the focus measure and the relevance of the image features estimated through the sharpening index:

$$\psi = \frac{1}{\Omega} \sum_{\forall k} \omega_k I_k, \tag{5.9}$$

where $\Omega = \sum_{k=1}^{K} \omega_k$ is a normalization factor and ω_k are weighting coefficients that replace the focus measure in (5.7) in order to allow an adaptive image fusion. The weigthing coefficients are computed based on the image content by applying a transfer function to the focus measure. This transfer function is in turn modulated by the sharpening index. Thus, the overall transformation will adapt to the image content. The definition of those coefficients is fully described below. For simplicity, pixel coordinates (x, y) have been omitted in (5.9). In the sequel, all symbols and equations refer to pixels at coordinates (x, y) unless otherwise is indicated.

The intensities in the AIF image corresponding to pixels that exhibit a strong visual pattern (nearly ideal behavior) are generated by giving a large weight to the intensities of those pixels with a high focus measure. In other words, the energy maximization scheme is held for pixels with a high sharpening index. In contrast, the intensities of the pixels that exhibit weak visual patterns and hence have a large noise influence are generated by averaging the original intensities over the whole focus sequence, thus giving preference to noise reduction. The objective of the transfer function is to provide a smooth continuous transition between these two extreme cases.

For illustration purposes, Fig.5.6 shows the desired behavior of the weights in (5.9) as a function of the normalized focus measure value for two different cases. Fig. 5.6a shows the weights for an idealized focus measure function $\varphi(x, y)$ that gives preference to those pixels with the highest focus values ($F_k \to 1$). In contrast, Fig. 5.6b shows the weights for a non-ideal focus function. In this case, pixels with a low focus value still have a significant contribution in the computation of the

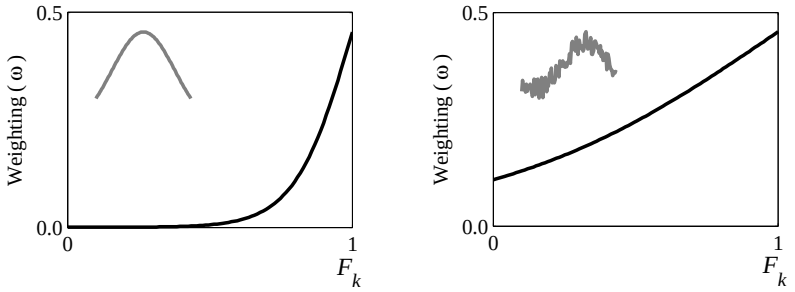


Figure 5.6: Selective weighting. (a) Weighting for pixels corresponding to a close-to-ideal focus function (high sharpening index). (b) Weighting for a noisy focus function (low sharpening index).

AIF image.

The behavior of ω in Fig. 5.6 is analogous to a high-pass filter in the frequency domain. In this work, this analogy has been exploited in order to propose a definition of ω. Digital filters are designed to cope with certain desired characteristics in the frequency domain while keeping an efficient time domain representation. Alternatively to traditional FIR and IIR filters, the hyperbolic tangent-based filters allow easy control of the cut-off frequency and the transition band (Wolberg, 1990; Basokur, 1998). A general pass-band hyperbolic tangent-based filter is defined in the frequency domain of f as :

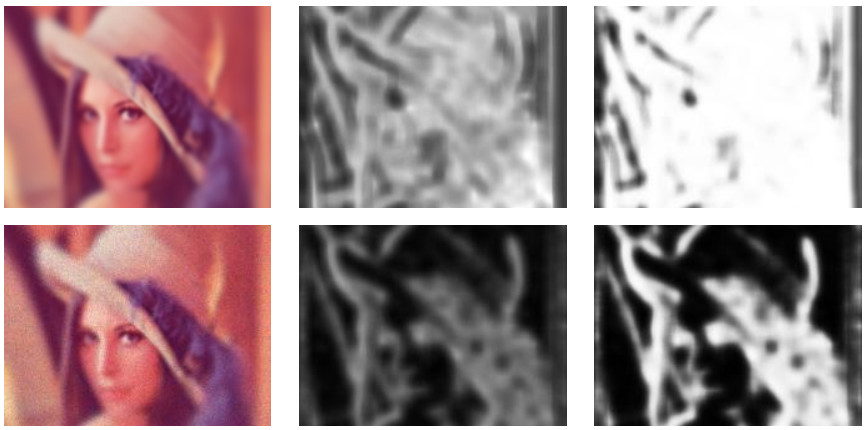
$$H(f) = \frac{\tanh(\phi(f \pm f_c)) + 1}{2}, \tag{5.10}$$

where f_c determines the band-pass frequency and ϕ controls the transition band (slope).

Equation (5.10) is suitable for the sought weighting in (5.9) since it has a fast exponential decay (e^{-f}) for points away from the cut-off frequencies and can be easily parameterized as a function of ϕ. In particular, since the normalized focus measure is between 0 and 1, ω is defined as a high-pass filter with a cut-off frequency of 1 (Basokur, 1998):

$$\omega(k) = \frac{\tanh(\phi(F_k - 1)) + 1}{2} \tag{5.11}$$

Equation (5.11) provides a continuous transition between the minimum and maximum values of F_k. The speed of that transition is modulated by the *sharpening index* ϕ which, in turn, is a function of the reliability as stated in the previous section.



(a) (b) (c)

Figure 5.7: Working principle of the proposed SAF algorithm. (a) Defocused frame. (b) Reliability measure $R(x, y)$. (c) Sharpening index $\phi_{x,y}$

Working principle

The key principle of the proposed *selective all-in-focus* (SAF) algorithm is its capability to adapt the image fusion process to both the image content and the amount of noise. For instance, Fig. 5.7a shows a frame from two synthetic sequences corresponding to the same scene but with different noise levels. The corresponding values of R and ϕ are shown in Fig. 5.7b and c, respectively.

In Fig. 5.7, the first row corresponds to a low-noise focus sequence. In this case, the R-measure is high all over the image, meaning that the pixel intensities and their corresponding focus measures are reliable. This leads to high values of the sharpening index. Thus, the image fusion process is mostly performed through an energy maximization scheme. In contrast, the second row of Fig. 5.7 corresponds to a sequence with high levels of noise. The sharpening index $\phi_{x,y}$ has high values only for those pixels where image features are strong enough to compensate for the effects of noise. In this case, the pixels corresponding to a high $\phi_{x,y}$ are fused by giving preference to high focus measure values (in analogy to a high-pass filtering). In contrast, the low values of $\phi_{x,y}$ in areas where the image patterns are weak will lead to a stronger smoothing, hence suppressing noise (in analogy to an all-pass fitering).

The main conceptual difference between the proposed approach and previous works is that, instead of applying the same smoothing rule to the whole image (by means of low pass filters, Gaussian pyramids or by removing wavelet coefficients), the image fusion is performed adaptively by taking into account the local features of the scene and the response of the focus operators to those features. This leads to a reduction of both noise and artifacts while preserving image texture.

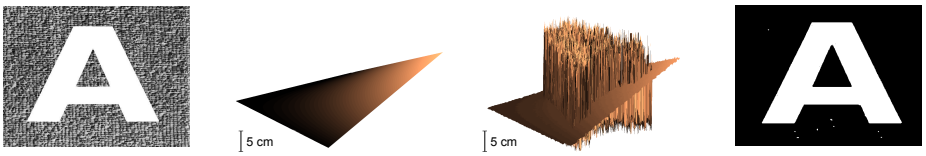


Figure 5.8: Binary reference mask. From left to right: all-in-focus image, ground truth, depth-map and binary reference mask.

5.5 Experiments and discussion

In this section, the proposed R-measure is applied for depth-map carving and focus stacking. For each application, the proposed approach is compared with state-of-the-art alternatives with both synthetical and real data sets.

5.5.1 Depth-map carving

In this section, the reliability measure is applied in order to determine the regions where shape-from-focus is unlikely to perform correctly. For each reconstructed sequence, the aim is to generate a binary segmentation mask that removes the pixels whose depth error is above a predefined tolerance.

Tests on simulated data

In synthetic sequences, the ground truth is accurately known. Therefore, the error in the depth-maps can be readily computed. For comparison purposes, a reference segmentation mask, $M_{ref}(x, y)$, is generated for each scene as:

$$M_{ref}(x,y) = \begin{cases} \text{true} & \text{if } |z(x,y) - z_G(x,y)| > e_T \\ \text{false} & \text{otherwise} \end{cases} \tag{5.12}$$

where $z(x, y)$ is the depth-map, $z_G(x, y)$ the corresponding ground-truth and e_T is the maximum error allowed for depth estimation (any pixel with a depth error higher than e_T is considered to be erroneous and discarded). Fig. 5.8 shows the generation of the reference mask for a synthetic scene. In this figure, pixels with erroneous depth estimates (error greater than e_T with respect to the ground truth) are marked as *true* (white) in the binary reference mask.

Once the reference mask is generated for every scene, a segmentation mask, $M(x, y)$, is obtained by applying a threshold to the reliability measure as described in section 5.3. In the ideal case, the segmentation masks should be equal to the reference mask so that all pixels whose depth estimation is incorrect are removed from the depth-map.

In order to generate the segmentation, 4-fold cross-validation has been applied for finding the reliability threshold, $R_{\min}$. The filtering quality is assessed by means of the accuracy, Acc:

$$Acc = 100 \sum_{(x,y)} \frac{C(x,y)}{N}, \tag{5.13}$$

where the numerator C is the number of coincident pixels between the actual segmentation mask and the reference one, and N is the total number of image pixels:

$$C(x,y) = \begin{cases} 1 & \text{if } M_{ref}(x,y) = M(x,y) \\ 0 & \text{otherwise} \end{cases} \tag{5.14}$$

In addition to (5.13), the precision (P) and recall (R) have been used:

$$P = \frac{tp}{tp + fp} \tag{5.15}$$

$$R = \frac{tp}{tp + fn}, \tag{5.16}$$

where tp and fp are the number of true positives and false positives, respectively, and fn is the number of false negatives of M with respect to M_{ref}, with an error threshold of $e_T = 5\%$ in (5.12).

Given the fact that texture is an important variable in the estimation of the focus level, it is straightforward to exploit the texture information of the scene as a cue for computing an alternative reliability measure. However, in order to apply a texture segmentation approach, an all-in-focus image of the focus sequence is required. This implies that this alternative can only be applied as a post-processing by assuming that an accurate all-in-focus image of the scene can be obtained. For comparison purposes, the all-in-focus image of each scene was computed using the software developed by Helicon Soft (2011) and has been fed into three alternative two-class texture classifiers.

A first texture classifier is simply obtained by applying the 24 Gabor filters described by Manjunath and Ma (1996) to the AIF image and then averaging the responses of all the filters for every pixel. Those filters are widely used for texture classification and segmentation. The average response of the filter bank is expected to yield high values in image regions with rich texture content and low values elsewhere. Thus, the average response is used to separate the image into two classes by simple applying a threshold. The threshold is selected by finding the value that yields the best classification rate in a training set. The second texture classifier is obtained by combining the responses of each individual Gabor filter using *Adaboost* (Freund and Schapire, 1995; Friedman et al., 2000). In this case, a *weak classifier* is generated by simply applying a threshold to the response of each

Table 5.1: Mean performance of different filtering methods using 8 simulated sequences with 4-fold cross-validation. Rank of each algorithm between brackets.

Method	Acc(%)	P(%)	R(%)
CAN	75.0 (4)	41.7 (4)	99.9 (1)
GAB	81.1 (3)	58.2 (3)	84.4 (4)
G+AD	95.7 (1)	83.4 (1)	91.3 (3)
DFIL	65.1 (5)	25.4 (5)	45.4 (5)
RBM	90.8 (2)	67.0 (2)	93.7 (2)

filter. The threshold that yields the best classification rate is selected. Adaboost is then used to combine each weak classifier in order to obtain the best classification rate in the training set.

In the experiments, a total of five depth-map carving approaches have been used in the comparisons: the Canny edge detector-based algorithm proposed by Muhammad and Choi (2011) (CAN), the mean response of the Gabor filters (GAB), the combination of Gabor filters using Adaboost (G+AD), the depth-map filtering-based algorithm proposed by Muhammad et al. (2009) (DFIL) and the proposed reliability-based method (RBM). In order to compare the effect of only the carving stage of these approaches, all the depth-maps have been computed with the traditional shape-from-focus framework without post-processing nor denoising (section 2.2.3).

The different filtering algorithms require tuning their own parameters by means of a training process. For each filtering algorithm, the training process was carried out using 4-fold cross-validation in a test set of 8 synthetical sequences generated using the algorithm described in appendix B. In each fold of the training process, the parameters of each filtering algorithm were adjusted in order to obtain the best classification rate in terms of accuracy.

Table 5.1 compares the mean performance of the different filtering methods. Fig. 5.9 shows a frame of the focus sequence and the filtered depth-maps using the evaluated algorithms for three out of eight sequences from the synthetic test set. In table 5.1, the cascade classifier based on Adaboost showed the best performance for the synthetic scenes in terms of accuracy and precision, whereas the proposed RBM shows the best performance in terms of recall.

Tests on real data

For complex real scenes, it is difficult to determine the ground truth accurately. In order to overcome this problem, the reference mask is manually generated by pre-computing the depth-map and marking those pixels whose estimated depth is incorrect. For a fair analysis of the results, a single reference mask was created for

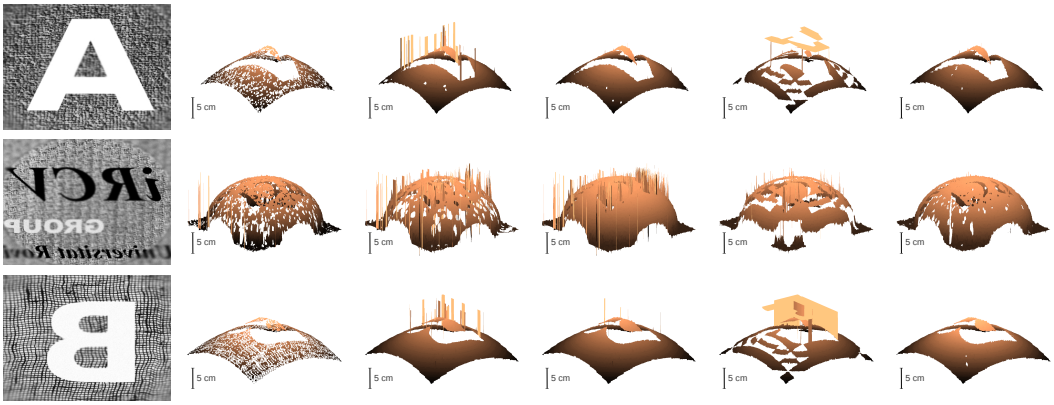


Figure 5.9: Performance comparison with synthetic sequences. From left to right: All-in-focus image, and depth-maps filtered with CAN, GAB, G+AD, DFIL and the proposed RBM, respectively.

Table 5.2: Mean performance of different filtering methods using 12 real sequences with 4-fold cross-validation. Rank of each algorithm between brackets.

Method	Acc(%)	P(%)	R(%)
CAN	61.0 (5)	41.6 (4)	99.7 (1)
GAB	78.1 (3)	56.8 (3)	83.6 (4)
G+AD	86.7 (2)	67.4 (2)	88.4 (3)
DFIL	63.8 (4)	38.3 (5)	27.1 (5)
RBM	88.4 (1)	68.1 (1)	93.4 (2)

each scene and used in all the experiments. Once the reference mask is constructed, the comparative tests are performed similarly to the synthetic sequences.

As shown in table 5.2, the proposed R-measure outperforms all the other methods in terms of accuracy and precision in the real sequences. In addition, it ranks between the first and second place in all the quality measures for both the synthetic and real sequences. Some methods, such as CAN, provide a high recall at the cost of low accuracy and precision. This behavior is illustrated in Fig. 5.9 and Fig. 5.10, in which some methods over-carve the depth-map, thus yielding a high recall at the cost of removing relevant information. Alternatively, some methods allow too much noise in the carved depth-map. In general, the proposed R-measure is a good tradeoff between the different quality measures by removing erroneous pixels while preserving the relevant information of the depth-map. In the experiments with real scenes, the reliability threshold, $R_{\min}$, had a little variation (between 12.3 and 15.3 dB). This is desirable since it suggests that the R-measure readily adapts to different imaging conditions and scenes without significant changes on its behavior.

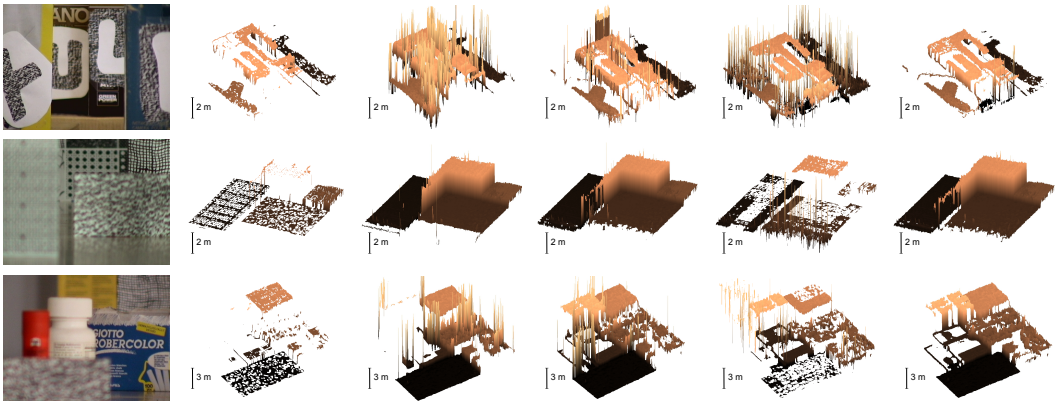


Figure 5.10: Performance comparison with real sequences. From left to right: All-in-focus image, and depth-maps filtered with CAN, GAB, G+AD, DFIL and the proposed RBM, respectively.

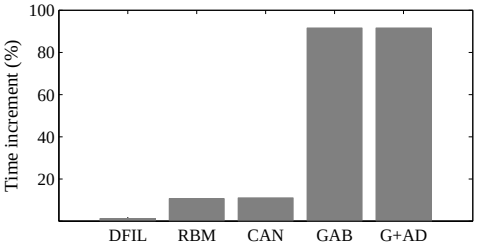


Figure 5.11: Computational cost of different filtering methods.

As previously stated, one of the main advantages of the proposed R-measure is that it efficiently integrates with focus-based techniques. The Matlab implementation of the classical SFF routine runs in approximately 4.05s for a sequence of 25 images of 640×640 pixels on an Intel 2 Quad processor at 2.5GHz and 4GB of RAM. Fig. 5.11 summarizes the time increment (as a percentage of the duration of the original SFF routine) for different depth-map carving alternatives. Notice that the computation time of the all-in-focus image for texture-based methods (CAN, GAB and G+AD) has not been included in Fig. 5.11 since a third-party software has been used for its computation. The simplicity and efficient integration of the proposed methodology into the shape-from-focus framework lowers the overall computational cost. According to Fig. 5.11, the computation of the R-measure yields a slight increase of approximately 10.6% in the computation time of the basic SFF stage.

5.5.2 Focus stacking

Several experiments have been conducted in order to assess the performance of the proposed SAF approach with different noise levels for real and synthetic focus sequences.

Tests on simulated data

The focus sequences synthetically generated as described in appendix B allow the availability of a ground truth for an objective estimation of the performance of the image fusion process. Based on a thorough review of the literature, five algorithms were selected for comparison:

1. *Helicon Focus*: an image fusion software produced by Helicon Soft (2011).
2. *Zerene Stacker*: a fusion software produced by Zerene Systems (2011).
3. Extended depth-of-field (EDF): a fusion algorithm based on wavelets proposed by Forster et al. (2004).
4. 3D extended depth of field (3D EDF): a fusion algorithm based on defocus modeling proposed by Aguet et al. (2008).
5. The algorithm proposed by Tian et al. (2011) based on the spatial frequency.

Figure 5.13 shows the mean performance of the different algorithms in terms of signal-to-noise ratio: $\text{SNR} = 20\log(\Sigma_{x,y} I(x,y)/\Sigma_{x,y}|I(x,y) - \psi(x,y)|)$ and the *universal quality index* (UQI) originally proposed by Wang and Bovik (2002). The i-th noise level corresponds to noise variances $\sigma_c^2 = \sigma_s^2 = 0.06i$. Fig. 5.12 shows a frame from four synthetic focus sequences, whereas Fig. 5.14 shows details of the all-in-focus images obtained from these sequences using the evaluated algorithms. For high noise levels, the difference in quality of the AIF image obtained with the different methods tends to increase in favor of the proposed SAF algorithm. The images shown in Fig. 5.14 correspond to the first noise level $(i = 1)$[2].

The results presented in this section show that the proposed method outperforms the other tested algorithms for synthetic sequences even at the lowest noise levels. For high noise levels, the results tend to evolve in favor of the proposed algorithm. For instance, in the first column of Fig. 5.14, the face has a smooth appearance, whereas highly-textured areas, such as the hat's feathers, are sharply recovered. In the image of the camera-man, the sky shows a clean appearance, whereas the contours of the man are well defined.

[2]A full resolution version of all the images shown in this work and the parameters of each stacking algorithm can be found online at http://www.sayonics.com/research/focus_fusion.html

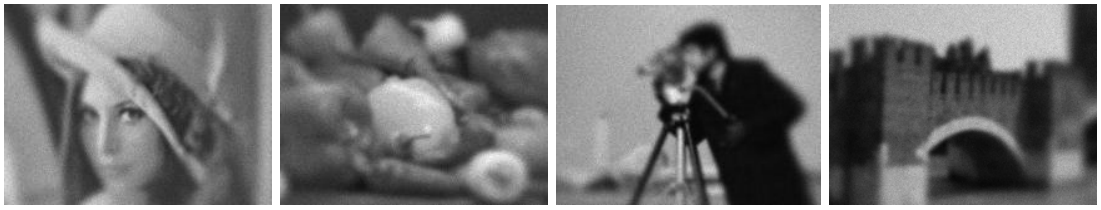


Figure 5.12: Synthetic focus sequences used for comparison. The images correspond to the first frame of the focus stack.

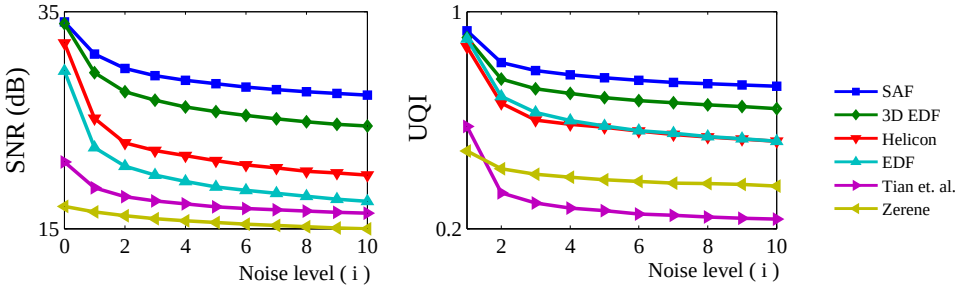


Figure 5.13: Performance comparison for synthetic focus sequences.

Tests on real data

Real focus sequences of 33 images of 640×480 pixels were acquired with a Sony SNC-RZ50P camera. In order to increase the noise level, the shutter speed was reduced. Since the intensity of an image pixel is proportional to the integration time (inversely related to the shutter speed), the loss of intensity is then compensated by the gain of the camera.[3] This leads to an increase of the *amplification noise*. The procedure for capturing real sequences with different noise levels can be summarized as follows:

1. The camera is adjusted to obtain the best-quality image of the scene and a focus sequence is captured. This sequence corresponds to noise level 0.

2. The gain of the camera is increased in order to raise the amplification noise. The shutter speed is increased in order to compensate for the illumination change. This sequence corresponds to noise level 1.

3. Since the camera gain and shutter speed can only be set to discrete predefined values, the histograms of the captured images must be equalized as necessary in order to keep constant illumination.

4. Steps 2 and 3 are repeated to compute the sequence corresponding to the i-th noise level.

[3]The lens aperture must remain unchanged in order to keep the same depth-of-field.

Figure 5.14: Image fusion for synthetic focus sequences obtained with different algorithms. From top to bottom: Zerene Stacker by Zerene Systems (2011), Algorithm of Tian et al. (2011), EDF (Forster et al., 2004), Helicon Soft (2011), 3D EDF (Aguet et al., 2008) and proposed SAF algorithm.

Table 5.3: Noise levels for real sequences

Noise level (i)	Gain [dB]	Shutter speed [s]
0	0	1/12
1	+16	1/75
2	+20	1/120
3	+22	1/150
4	+26	1/215
5	+28	1/300

Figure 5.15: Real scene at increasing noise levels. Left to right: 0th, 1st, 3rd and 5th noise level.

Table 5.3 summarizes the camera configuration used for the acquisition of real sequences at different noise levels. Fig. 5.15 shows a frame of a particular scene with different noise levels.

Tests have been conducted on both color and gray scale images. For the color images, each frame was converted to gray scale in order to compute both the focus measure and the selectivity index. Then, the fusion rule in (5.9) was independently applied to each color channel.

For the real sequences, it is not possible to compute an objective quantitative performance measure since the ground-truth is not available and the quality of the results must be subjectively determined by simple observation. However, it is possible to assess the impact of noise over the fusion process by comparing the AIF image obtained from a noisy sequence against the AIF image obtained from the sequence with the lowest noise level. Thus, in Fig. 5.16, the SNR is computed using the all-in-focus image obtained from the sequence with the lowest noise level as a reference. For an algorithm to be robust to noise, the AIF images of sequences with high noise will be less corrupted and will, therefore, have a high SNR. Fig. 5.17 shows details of the all-in-focus images obtained from a color focus sequence and a gray-scale sequence using Helicon Focus, 3D EDF and the proposed SAF algorithm.

In real sequences, the results show that the SAF algorithm is the least sensitive to noise, confirming the results obtained with synthetic sequences. For example, in the detail images of Fig. 5.17, the background always presents a cleaner ap-

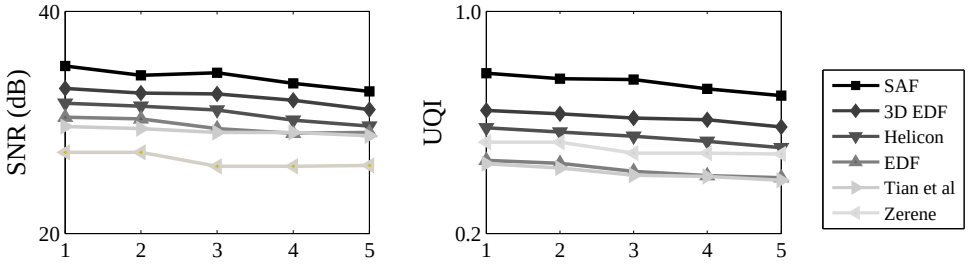


Figure 5.16: Performance comparison for real focus sequences. x-axis: noise level (i).

(a) (b)

Figure 5.17: Image fusion for real focus sequences obtained with different algorithms. (a) For color focus sequence. (b) For gray-scale focus sequence. From top to down: first row, Helicon Focus software by Helicon Soft (2011); second row, 3D EDF algorithm (Aguet et al., 2008); third row, proposed SAF algorithm.

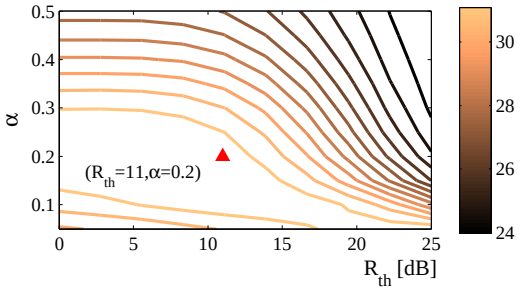


Figure 5.18: Level plot of the mean SNR of synthetic sequences as a function of the parameters α and S_{th}. The best performance is obtained for $S_{th} = 11$ and $\alpha = 0.2$

.

pearance, whereas edges, letters and contours are sharply defined even for thin and small characters. However, a white halo can be observed near the edges of the white letters within the detail image in Fig. 5.17. This effect is barely noticeable and is observed at high noise levels around bright white areas surrounded by a dark background. This halo can be due to the fact that bright spots have a larger spread than dark ones when defocusing, and the radiance of these spots may "leak" into dark areas during the image fusion process.

Algorithm's parameters

As shown in section 5.4.1, the proposed algorithm depends on two parameters. The values of those parameters used in the results shown in this work for both synthetic and real sequences correspond to $R_{th} = 11$ dB and $\alpha = 0.2$. These parameters have been selected experimentally using the synthetic sequences and recording the mean SNR for each parameter pair.

As shown in Fig. 5.18, the parameters were selected in order to maximize the performance for the synthetic sequences (in terms of SNR). In order to assess the sensitivity of the proposed approach to these parameters, the synthetic focus sequences corresponding to noise level 1 were processed with variations of those parameters of $\pm 15\%$. The maximum variation observed in the SNR was -1 dB (3.1%). With this variation, the proposed focus stacking algorithm still outperforms the closest competing algorithm (3D EDF). The variation of SNR for real sequences was 1.5%. This results show that the proposed method is reasonably insensitive to its parametrization.

The need for parameters is common in the evaluated AIF algorithms. Table 5.4 summarizes the parameters of the all-in-focus algorithms compared in the experiments. The fact that the transfer function of (5.11) is defined in terms of sigmoids guarantees that the intensity of a given pixel of the AIF image will always be a combination of the pixel intensities of the focus sequence. This has a positive

Table 5.4: Parameters of the compared focus stacking algorithms.

Algorithm	Configuration
SAF (proposed)	$R_{th} = 11$ [dB], $\alpha = 0.2$.
3D EDF	$n_0 = 0.2$, and $n_1 = 1.3$
Helicon Focus	Method: B, radius: 8, smoothing: 4.
EDF	Complex wavelets, Daubechies 6; scales: 8; denoising: 10%.
Zerene Stacker	Method *DMax* (with defaults)
Tian's	Average blur: 5, $\alpha = 1$, $\beta = 0.5$.

impact on the stability of the algorithm.

In terms of computational cost, the least complex all-in-focus methods are those based on the spatial frequency. These methods usually imply the application of a focus measure to each image of the focus sequence, followed by a fusion rule. In the second place, the pyramid-based and wavelet-based approaches usually require a forward transform, a combination step applied to the obtained sub-images or sub-bands, and an inverse transform. The cost of the forward and inverse transforms increases with the number of levels of the pyramid. The methods with the highest computational costs are those based on defocus modeling (e.g., 3D EDF).

The different methods compared in this work were obtained from different sources and platforms (e.g., Java, Matlab, C). Therefore, an objective quantitative comparison in terms of computation time is not provided. Notwithstanding, the efficiency of the proposed approach is between that of spatial-based methods and wavelet-based methods. Similarly to spatial-based methods, the SAF algorithm requires the application of a focus measure followed by a fusion rule. However, the computation of the selectivity measure represents an additional cost. In spite of that, the computation of both the focus measure and the selectivity measure is simple and fast. In particular, the computational complexity of the proposed algorithm is $\mathrm{O}(NKh^2)$, where K is the number of images, N the number of pixels in each image and h the size of the support window of the focus measure operator. The Matlab implementation of the proposed algorithm fuses a sequence of 50 gray-scale images of 640×480 pixels in approximately 7.0 s running on an Intel 2 Quad processor at 2.5 GHz and 4GB of RAM.

5.6 Summary

The reliability measure, R, presented in this chapter is aimed at predicting the confidence of the estimated focus measure corresponding to each image pixel. The R-measure takes advantage of the theoretical focus profile model developed in the

previous chapter in order to assess how close to the ideal behavior the estimated focus measure in a real image sequence is. A straightforward application of the R-measure in depth-estimation consists of detecting pixels whose reliability is below a given threshold in order to discard them while preserving the useful information of the depth-map of complex scenes. The results presented in section 5.5.1 show the advantages of the R-measure with respect to different alternatives.

Texture-based methods (e.g., CAN, GAB and G+AD) base their response on the texture information of the all-in-focus images with an important drawback: the all-in-focus image does not take into account the variations of the focus function along the z-axis (as a function of focus), which is an important factor in the performance of SFF. These variations are mainly due to CCD noise or optical effects such as the curvature field, image shift or artifacts. This could explain why they perform better only in the ideal case (in the synthetic sequences). In contrast, the proposed R-measure performs satisfactorily in both synthetic and real sequences.

Improving the robustness of shape-from-focus to different imaging factors and acquisition conditions is fundamental for enhancing the quality of the obtained depth-maps. Complementarily, the proposed R-measure is aimed at predicting the performance of the depth estimation in order to take the most advantage of the shape-from-focus technique. In addition to depth estimation, the proposed R-measure has been exploited in focus stacking for the generation of all-in-focus images robust to noise.

From the results obtained for a same scene with different noise levels, it is evident that the proposed approach responds to the image content selectively. Therefore, as the noise level increases, the fusion process provides a smooth low-noise response in areas where the focus function presents low PSNR, while reducing the negative impact on image features.

CHAPTER 6

Shape estimation from autofocus

Classical focus-related techniques, such as autofocus, focus stacking, shape-from-focus, shape-from-defocus and focus calibration, exploit the knowledge about different parameters of the acquisition process and the focus sequence in order to obtain additional information of the imaged scene. For instance, the focus of the camera and the estimated relative focus measure of every image pixel is exploited in shape-from-focus/defocus for 3D scene reconstruction. In turn, the knowledge about the current and previous focus positions are exploited by search strategies in order to speed up autofocus. In this chapter, a more challenging scenario is explored. A sequence of images corresponding to an autofocus sequence is processed in order to infer geometric information of the scene without knowledge about the current configuration of the camera.

Section 6.1 introduces the proposed approach, namely *shape from autofocus* (SFA). A fundamental concept for the proposed approach, the *focus signal*, is introduced in section 6.2. The methodology for exploiting autofocus in order to obtain information about the scene geometry is presented in section 6.3. The proposed approach is experimentally evaluated in section 6.4. The obtained results are finally summarized in section 6.5.

6.1 Introduction

As stated in previous chapters, most digital cameras currently have an autofocusing mechanism aimed at adjusting the mechanical configuration of the lens-sensor system in order to capture sharp images without human intervention. Depending on both the acquisition device and the imaged scene, the autofocus stage may take from a fraction of a second up to several seconds. The duration of this process, which takes place before each image is captured, is often seen as an undesirable effect that should be minimized in order to speed up the acquisition process.

Alternatively, this chapter proposes a new interpretation for the autofocusing process by exploiting the implicit information about the scene geometry found in the variations of the focus level. The claim is that, being autofocus an unavoidable part of systems with limited depth-of-field, it is possible to take advantage of this process in order to infer useful information about the imaged scene. Based on the results obtained in previous chapters, autofocus is modeled as a time-variant interaction between the capturing device and the observed scene, showing that each imaged point generates a pattern, or *focus signal*, that mainly depends on the configuration of the lens-camera system and the scene geometry. Since at every instant, the lens-camera system has the same configuration for all imaged points, the scene geometry (in particular its approximate depth-map) can be estimated as a function of the different *focus signals*, where the focus signals are the focus measure values as a function of time.

Fig. 6.1 shows an image stack of a video sequence recorded while a camera is autofocusing on a real scene. The scene is divided into a discrete number of regions of interest. A coarse depth-map of the scene is recovered by clustering the focus signals extracted from each region of interest. The coarse approximation of the *signal clusters* depicted in Fig. 6.1 can be interpreted as a segmentation process through which the image is segmented into disjoint regions by taking into account the geometry of the scene rather than the color or texture features typically used in segmentation tasks.

To the best of our knowledge, the problem inferring geometrical information of the imaged scene by processing an unordered set of focus samples has not been tackled previously. In contrast to shape-from-focus and shape-from-defocus approaches, the proposed technique does not require an accurately controlled -and therefore slow- acquisition process, nor previous knowledge or calibration of the parameters of the acquisition device. In addition, the proposed approach does not rely on the estimation of the maximum focus value (as in shape-from-focus) or the relative degree of focus (as in shape-from-defocus), which are sensitive to noise and depend on the image content. Alternatively, the whole set of focus measure values are used, thus yielding more robust results.

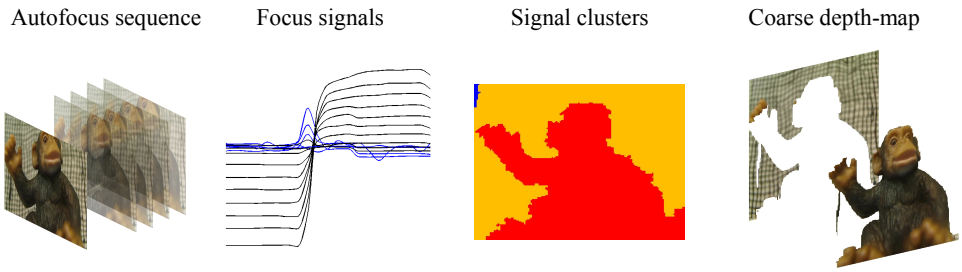


Figure 6.1: Coarse depth-map estimation from autofocus sequence. A coarse depth-map is generated by clustering the focus signals extracted from the image frames of an autofocus sequence.

6.2 Focus signal model

The focus signal is defined as the focus measure value as a function of time, φ vs. t, corresponding to a particular image pixel (or image region). As previously demonstrated in chapter 4, the focus level is a function of the camera constant, κ, the target position, u_x and the focus of the camera itself, u. As a result, by assuming that the only parameter of the acquisition device that changes during autofocusing is the camera focus u, the focus signal can then be expressed as:

$$\varphi(t) = \varphi(u_x, u(t)), \tag{6.1}$$

where $x(\cdot)$ is an unknown function and $u(t)$ is the variation of the camera's focus as a function of time.

Since the focus of the camera, $u(t)$, is the same for all the scene points at each time instant, objects at different positions can be univocally distinguished from their corresponding focus signals. In other words, two objects at different target positions, u_{x1} and u_{x2}, will yield different focus signals, $\varphi_1(t) \neq \varphi_2(t)$. This is a consequence of the one-to-one correspondence between the focus level and the scene geometry previously established in equations (4.7) and (4.29). This particular approach has the advantage of not relying on the absolute or relative degree of focus (which, in turn, depends on the image content, image noise and the focus measure operator) and requires no knowledge about the camera or the acquisition parameters. This fact is illustrated in the following experiment: Fig. 6.2a shows a highly-textured target in front of a camera at a fixed distance, u_{x1}. An image sequence is then captured by moving the focus of the camera back and forward between +0.5m and -0.5m around the target position, thus covering a range of 1m. Fig. 6.2b shows the in-focus position as a function of time. The experiment is then repeated by moving the target to a new position, u_{x2}. Fig. 6.2c plots the focus signals for the first experiment (φ_1) and the second experiment (φ_2).

Notice that for both target positions in the previous experiments, the informa-

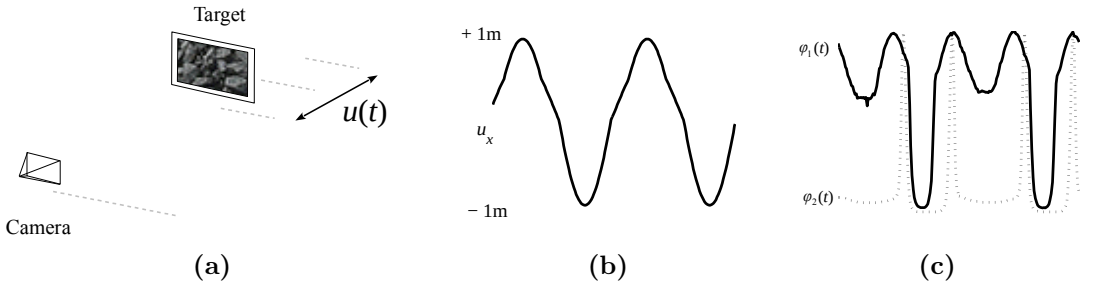


Figure 6.2: Focus signal as a function of the target position. (a) Test setting. (b) Focus trajectory (the focus of the camera as a function of time). (c) Focus signals corresponding to the same target at two different positions.

tion of the focus position is lost since the focus signals are a function of time. The parameters of the camera, namely the lens focal length, aperture and pixel size, are assumed to be unknown. In addition, since the only difference between both targets is their distance from the camera, they cannot be differentiated by means of the focus level or image content (color, texture, etc). Remarkably, it is clear from Fig. 6.2c that targets at different positions clearly yield different focus signals. Notwithstanding, this is only valid for the ideal noiseless, aberration-free and highly-textured case. Further considerations must be taken into account in order to apply it in real imaging conditions, since the effects of noise, camera movement during autofocusing, vibration, among others, make it harder to distinguish the focus signals corresponding to different objects.

Let us consider the simplest case of a fronto planar object at a distance u_x from the camera, partitioned into a finite set of N non-overlapping regions of interest (ROI): $\{\Omega_n | n = 1, 2, \cdots, N\}$. In the limit case, when the area of each ROI tends to 0, each Ω_n corresponds to an image pixel. The focus signal corresponding to the n-th ROI, $\mathbf{x}_n$, is defined as the sequence of values of the focus measure over Ω_n at subsequent time instants. From this perspective, the autofocusing process yields a set of N time-varying focus signals of finite duration. Specifically, for the case of discrete-time signals, the autofocusing process yields a set of N vectors (signals) such that the n-th vector has τ elements, $\mathbf{x}_n = [\varphi_n(t_1), \varphi_n(t_2), \cdots \varphi_n(t_\tau)]^T$, with τ being the number of frames (i.e., time instants) of the autofocus sequence.

In real imaging conditions, Subbarao and Tian (1998) showed that the expected value μ of the focus measure φ in a real noisy image is the sum of the focus measure of the ideal noiseless image plus a constant value proportional to the noise variance. According to the imaging chain previously presented in section 2.1.2, noise is independent of the lens position. It actually depends on the camera's sensor. Therefore, two observed focus measures, $\varphi(t_1)$ and $\varphi(t_2)$, corresponding to different lens configurations are statistically independent random variables (Subbarao and

Tian, 1998; Tsai and Chen, 2012). In this work, it is further assumed that these variables are normally distributed. Therefore, each focus measure value can be treated as a random variable with a Gaussian probability density function (PDF). Thus, as long as the focus signals correspond to ROIs at the same distance from the camera, they can be considered to be samples from a stochastic Gaussian process whose probability density function can be modeled as a multivariate normal distribution (Yates and Goodman, 1999): $\mathbf{x} \sim \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M\}$, where:

$$\mathcal{N}(\mathbf{x}; \mu, \boldsymbol{\Sigma}) = (2\pi)^{-\tau/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \tag{6.2}$$

The mean $\boldsymbol{\mu}$ is the vector $[\mu_1, \mu_2, ..., \mu_\tau]^T$ corresponding to the *ensemble* average, that is, the average of all the focus signals for each time instant, and $\boldsymbol{\Sigma}$ is a diagonal correlation matrix: $diag(\sigma_1, \sigma_2, ..., \sigma_\tau)$. The diagonality of $\boldsymbol{\Sigma}$ comes from the statistical independence of each $\varphi_m(t)$.

6.3 Shape from autofocus

The algorithm to retrieve the coarse depth-map of a scene from the images captured during autofocus can be divided into three stages: first, an initial scene segmentation is obtained by clustering the focus signals according to the formulated model. Then, a refinement step is carried out in order to improve the initial segmentation, yielding a coarse approximation of the imaged scene. Finally, the boundaries of the coarse approximation are improved by incorporating the information obtained from classical image segmentation approaches. These stages are described below.

6.3.1 Signal clustering

Equation (6.2) corresponds to the PDF that models the focus signals originated from a single planar object perpendicular to the camera. However, in complex scenes with many different objects at different positions, a more flexible model is required. For this purpose, a *Gaussian mixture model* (GMM) is introduced. GMMs are a popular tool with application to statistical signal processing (McLachlan and Peel, 2000), speech recognition (Wrigley et al., 2005) and biomedical signal processing (Wang et al., 2011), among many others.

In particular, let us partition the scene into N non-overlapping ROIs, Ω_1, Ω_2, $\cdots$, Ω_N, of fixed size. In this work, ROI sizes greater than one pixel are used for two reasons: first, to provide a high SNR for a reliable measurement of the focus level and, second, to reduce the number of sample data, thus improving the processing

time. The sample data X is the set of the N focus signals corresponding to the N ROIs: $X = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$. The aim is to find a probability density function $\Gamma(\mathbf{x}, \theta)$, $\mathbf{x} \in X$, corresponding to a family of multivariate Gaussian distributions that is most likely to have generated the sample data:

$$\Gamma(\mathbf{x}, \theta) = \sum_{m=1}^{M} \omega_m \mathcal{N}_m(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \tag{6.3}$$

where M is the number of Gaussians in the model, $\mathcal{N}(\cdot)$ is defined as in (6.2), ω_m is the weight of the m-th Gaussian function (mixing probability) and θ is the set of parameters of the model: $\theta = \{\omega_m, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m | m = 1, 2, ..., M\}$. Since $\Gamma(\cdot)$ is a probability density function, the weights ω_m must add to one: $\sum_{m=1}^{M} \omega_m = 1$. For the sake of simplicity, $\mathcal{N}_m(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$ will be abbreviated as $\mathcal{N}_m(\mathbf{x})$ in the sequel.

The problem now reduces to finding the parameter set $\hat{\theta}$ that maximizes the data log-likelihood:

$$\hat{\theta} = \arg\max_{\theta} \{\log P(X|\theta)\}, \tag{6.4}$$

where, $P(X|\theta) = \prod_{n=1}^{N} \Gamma(\mathbf{x}_n, \theta)$.

There is no closed-form solution for (6.4), although it is differentiable. Therefore, any general purpose non-linear optimizer can be used to solve it. Notwithstanding, the *expectation-maximization* (EM) algorithm provides a convenient solution for the case of Gaussian mixtures. The EM algorithm iteratively generates a sequence of estimates, θ^k, from an initial guess θ^0. In the literature, the initial parameters are often calculated by applying a supervised clustering algorithm to the data (e.g., k-means) when the number of Gaussians is known. In this work, the *x-means* algorithm (Pelleg and Moore, 2000) has been used for automatically finding the number of clusters M and initializing the parameter vector. A detailed description of the EM algorithm for Gaussian mixtures can be found in (McLachlan and Peel, 2000; Nabney, 2001). The relevant details are summarized below.

The EM algorithm has two steps, namely expectation step (E-step) and maximization step (M-step). Let $P(m|\mathbf{x_n})$ denote the probability that the n-th sample point corresponds to the m-th mixture in (6.3). At the k-th iteration, the E-step corresponds to:

$$P^{(k+1)}(m|\mathbf{x}_n) = \frac{\omega_m \mathcal{N}_m(\mathbf{x}_n)}{\sum_{m=1}^{M} \omega_m \mathcal{N}_m(\mathbf{x}_n)} \tag{6.5}$$

The M-Step requires the global optimization of the parameter set as (Nabney,

2001):

$$\omega_m^{(k+1)} = \frac{1}{N}\sum_{n=1}^{N} P^{(k)}(m|\mathbf{x}_n) \tag{6.6}$$

$$\boldsymbol{\mu}_m^{(k+1)} = \frac{1}{N\omega_m^{(k+1)}}\left(\sum_{n=1}^{N} P^{(k)}(m|\mathbf{x}_n)\mathbf{x}_n\right) \tag{6.7}$$

$$\sigma_{m,n}^{(k+1)} = \frac{1}{\tau N\omega_m^{(k+1)}}\left(\sum_{n=1}^{N} P^{k}(m|\mathbf{x}_n)(x_{m,n}-\mu_{m,n}^{(k+1)})^2\right)^{-1/2}, \tag{6.8}$$

where $\sigma_{m,n}$ is the n-th element of the diagonal of $\boldsymbol{\Sigma}_m$.

Once the model in (6.3) has been found, the sample data can be clustered by assigning each focus signal to the Gaussian with the highest posterior probability for that signal. Thus, the n-th focus signal (and its corresponding ROI Ω_n) will be assigned to cluster C_n according to the following rule:

$$C_n = \arg\max_m P\left(\mathbf{x}_n|\omega_m\mathcal{N}_m(\mathbf{x})\right) \tag{6.9}$$

Equation (6.9) merges the ROIs that likely correspond to the same Gaussian of the mixture.

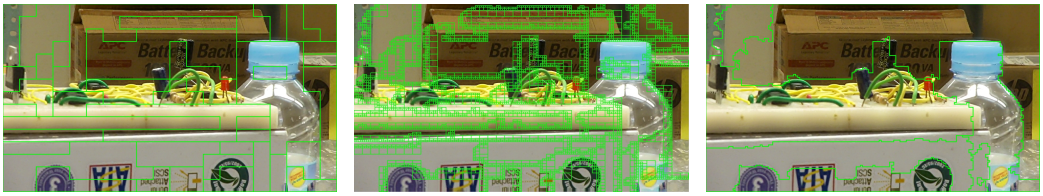
6.3.2 Scene refinement

By dividing the scene into ROIs of finite size greater than one pixel, the robustness of the focus signals is improved at the cost of spatial resolution. Therefore, the clustering described in the previous section corresponds to an initial segmentation that must be further refined.

The initial segmentation is obtained by only using the information in the focus signals without any assumptions about the spatial relationship with the scene ROIs. In order to refine the boundaries between regions, a quad-tree subdivision is proposed. The latter is carried out by dividing each sub-window belonging to the boundaries among two or more different regions into four quadrants of equal size (see Fig. 6.3a). For each quadrant $\{Q_i \subset \Omega_n | i = 1, 2, 3, 4\}$ a focus signal $\mathbf{x}_i$ is computed using the corresponding scene region. Each quadrant Q_i is then compared against the clusters in its neighborhood and reassigned to the cluster C_n with the highest likelihood:

$$C_n = \arg\max_{j\in Nh(Q_i)} P(\mathbf{x}_i|\omega_j\mathcal{N}_j(\mathbf{x})), \tag{6.10}$$

where $Nh(Q_i)$ is the set of clusters in the neighborhood of Q_i.



(a) (b) (c)

Figure 6.3: Scene refinement. (a) Initial clustering. (b) Quad-tree subdivision: the ROIs belonging to the boundaries of two or more clusters are split using a quad-tree partitioning. (c) Final clusters

The quad-tree subdivision process is recursively iterated until the smallest integer ROI size is reached. The re-assignment rule of (6.10) is parameter-free and consistent with the model developed in the previous section. In order to compensate for variations in the focus measure due to changes in illumination and the amount of texture in the scenes, the focus signals are standardized to have zero mean and a standard deviation of one.

In the first stage (signal clustering), the number of Gaussians has been selected using x-means, a general purpose unsupervised clustering algorithm not designed for this particular application. The non-optimal selection of the number of components of the GMM leads to an over-segmented scene. Therefore, a consistency test is applied in order to fuse neighboring clusters that are unlikely to correspond to different objects. In order to be consistent with the proposed autofocus model, that fusion should take into account the similarity between the distributions that correspond to each cluster (e.g., the Kullback-Leibler divergence for the case of univariate distributions). However, there is not such a similarity measure for the multivariate case. Alternatively, in this work, two clusters a and b are fused if they satisfy:

$$||\boldsymbol{\mu}_a - \boldsymbol{\mu}_b|| < \min\left\{||\boldsymbol{\sigma}_a||, ||\boldsymbol{\sigma}_b||\right\}, \tag{6.11}$$

where $\boldsymbol{\sigma}_a$ and $\boldsymbol{\sigma}_b$ are vectors of the form $[\sigma_1, ..., \sigma_\tau]^T$ corresponding to the traces of $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$, respectively.

Equation (6.11) aims at reducing the number of clusters (over-segmentation) and merges two clusters whenever the inter-mean distance lies within a hypersphere of radius $||\boldsymbol{\sigma}||$. The scene refinement in a real scene is illustrated in Fig. 6.3. Fig. 6.3a shows the initial segmentation obtained by clustering the focus signals using (6.9). The cluster's boundaries are further refined using (6.10), as shown in Fig. 6.3b. The final segmentation shown in Fig. 6.3c is obtained by merging similar clusters by means of (6.11).

6.3.3 Post-processing

The coarse approximation of the scene can be interpreted as an object-oriented scene segmentation: that is, the scene is segmented according to geometrically discontinuous objects that yield different focus signals. On the one hand, this representation of the scene is quite meaningful in the sense that it provides a simple description of the real geometry of the scene by clustering objects according to their depth and spatial location with respect to the camera. On the other hand, the coarseness of the obtained segmentation - more specifically, its contours- is an undesired artifact that should be minimized in order to improve the accuracy of the geometric description of the scene. However, it is important to remark that these results have been obtained exclusively based on the information extracted from the focus signals, with classical image segmentation cues, such as color, texture, local brightness and edges not having been explicitly exploited. As a result, the coarse reconstruction can be further improved by incorporating the information of these cues as shown below.

Image segmentation (contour detection and region merging) is a fundamental research field in computer vision. Intensive efforts have been devoted in order to produce more accurate and faster algorithms under the challenging scenario of real scenes. The problems and contributions in this particular field are not reviewed in this dissertation. For a detailed review of this research field, the reader is referred to the work by Arbelaez et al. (2011); Melendez (2010) and Serrano (2010).

This stage takes advantage of the state-of-the-art in image segmentation in order to post-process and improve the coarse depth-map obtained by the methodology described in the previous sections. In particular, the statistical region merging approach proposed by Nock and Nielsen (2004) has been selected due to a combination of reasonable performance, efficiency, publicly available implementations and low parametrization. The proposed segmentation-driven post-processing is simply carried out in two steps as described below.

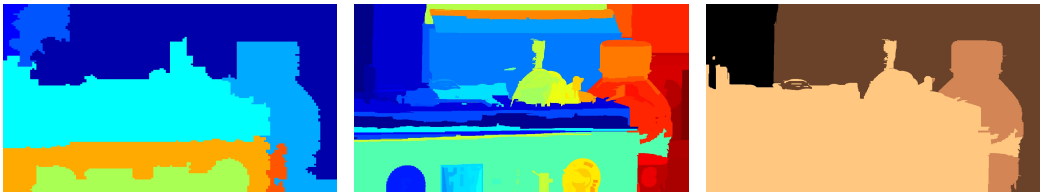
1. The imaged scene is divided into P super-pixels $\{\chi_p | p = 1, 2, \cdots, P\}$ using a segmentation algorithm (e.g., statistical region merging).

2. Each generated super-pixel, χ_p, is assigned to a cluster of the coarse approximation of the scene. In particular, a sub-region χ_r is assigned to the $\tilde{n}$-th cluster that maximizes the overlap:

$$\tilde{n} = \arg\max_{n} \sum_{(i,j)} \mathcal{O}_n(i,j), \tag{6.12}$$

where,

$$\mathcal{O}_n(x,y) = \begin{cases} 1 & \text{if } \chi_p(x,y) \in C_n \\ 0 & \text{otherwise} \end{cases} \tag{6.13}$$



(a) (b) (c)

Figure 6.4: Segmentation-based post-processing. (a) Coarse approximation. (b) Segmented image corresponding to the scene shown in Fig. 6.3. (c) Post-processed scene.

The image segmentation-based post-processing is illustrated in Fig. 6.4. From Fig. 6.4b it is evident that the image segmentation algorithm yields an over-segmented result, arguably due to the complexity of the particular scene. In contrast, the post-processed result shown in Fig. 6.4c is more accurate in terms of the real scene geometry. It is important to remark that the cooperative integration of different depth cues is an important problem in vision research (Atkins et al., 2000). In this scope, the existing approaches have mostly dealt with the integration of low- and high-level perceptual cues not related to focus, such as contours, optical flow and image features (DeCarlo, 2002; Triesch et al., 2002).

Without any prior knowledge about the parameters of the camera nor the lens configuration, the proposed algorithm is not capable of estimating the absolute position of the objects in the scene. Notwithstanding, it is possible to retrieve information about their *relative* position by finding the maximum of the mean focus signals and measuring the time separation between the peaks corresponding to different clusters. Even for a lens movement of unknown speed and direction, the time separation of the focus peaks is proportional to the real distance of the objects. The location of the focus peak is based on the same principle exploited in SFF for depth estimation. However, the focus measure is computed in this case by using a large support region instead of a small pixel neighborhood. This yields an increased robustness.

6.4 Experiments and discussion

The proposed method for obtaining and estimating the coarse depth-map of a scene is evaluated in this section. Test sequences have been obtained through different off-the-shelf cameras: a Logitech Orbit AF (*webcam*) and a compact digital photographic camera Sony DSC-HX5 (*photocam*). The autofocus sequences were obtained by simply pointing each camera to the scene and activating the autofocus mechanism. The autofocus sequence was recorded on video. Each recorded

Table 6.1: Real test sequences.

Sequence	Source	Size (pixels)	No. frames	Duration (s)
1	photocam	1280×720	9	0.30 @30 fps
2	webcam	800×600	10	0.40 @25 fps
3	webcam	640×480	14	0.56 @25 fps

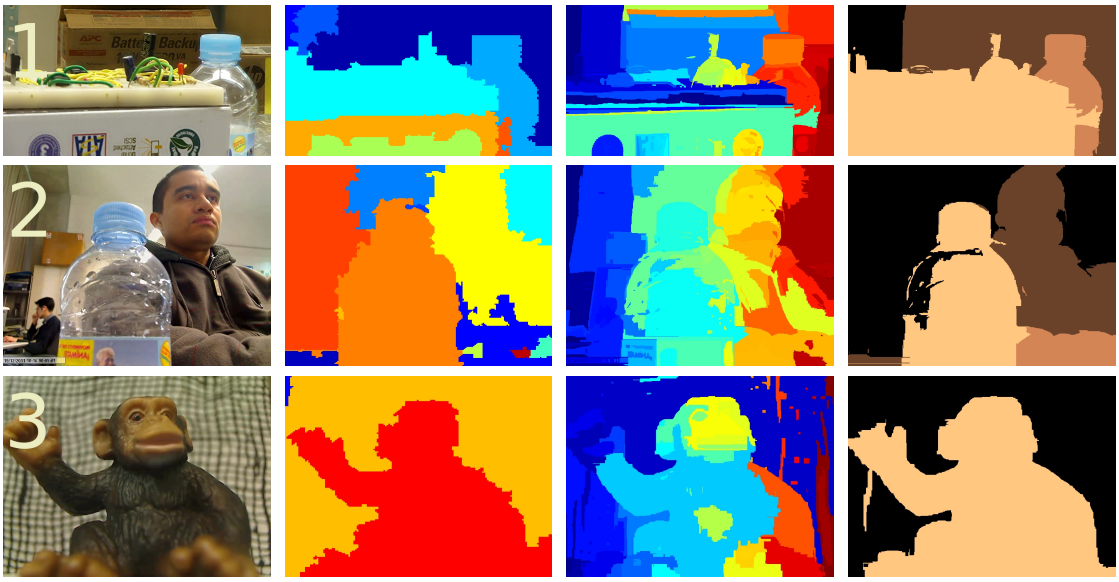


Figure 6.5: Depth-maps estimated using the proposed method for different image sequences. (a) All-in-focus image of each scene. (b) Generated clusters without post-processing. (c) Scene segmentation using the statistical image segmentation proposed by Nock and Nielsen (2004).(c) Obtained depth-map after post-processing. The lighter the color the closer the object.

sequence was then converted to separate image frames in order to process them. In the following experiments, the image sequences are identified with a number on the top left corner. Details about the capturing device, number of frames, image size and approximate duration of each sequence are summarized in table 6.1.

6.4.1 Coarse depth-map generation

This section describes the results obtained using the proposed SFA approach for the estimation of coarse depth-maps from the image frames of real autofocus sequences. Each row of Fig. 6.5 corresponds to a different image sequence. These sequences have been selected since they represent challenging scenarios with different imaging and acquisition conditions.

In particular, the first row of Fig. 6.5 corresponds to a video sequence of ap-

proximately 0.3s captured with a photographic camera. This particular scene has several challenging features for both classical segmentation approaches and focus-based depth estimation techniques, such as highly-textured and weakly-textured areas, objects with multiple colors and texture patterns, low-contrast regions, and reflective surfaces, among others. The coarse segmentation obtained using the proposed signal clustering approach without post-processing (second column) provides a meaningful description of the scene in terms of the scene's geometry. Thus, the two foreground objects are clearly isolated from the background. However, the objects' boundaries are coarse and inaccurate. Conversely, the segmentation shown in the third column of this figure provides an accurate description of the scene in terms of region boundaries and both color and texture consistency. However, the regions of the segmented image are difficult to interpret in terms of the scene geometry. The coarse depth-map (last column) shows a meaningful description of the scene's geometry with improved region boundaries. Different regions corresponding to the same object are successfully merged despite the differences in texture and color content. In addition, the depth-map provides useful information about the objects' depth ordering and layout in the scene.

The second and third rows of Fig. 6.5 correspond to image sequences captured with a webcam. In addition to the complexity of the imaged scene, the particular characteristics of the acquisition device, such as vignetting, distortion and image shift, pose an additional challenge. Notwithstanding, the obtained coarse depth-maps are meaningful and accurate in terms of both the region boundaries and the layout of the objects with respect to the camera.

The inherent complexity of the analyzed scenes is well illustrated in the segmentations shown in the third row of Fig. 6.5, where single objects are segmented in multiple regions due to differences in texture and color patterns. In contrast, the proposed algorithm yields simple and meaningful descriptions of the analyzed scenes for a wide variety of imaging conditions, as shown in the last column of Fig. 6.5. The depth-maps estimated with the proposed approach are rather simple but allow for clearly identifying objects at different distances from the camera.

In contrast to previous approaches, the proposed method does not require:

1. Knowledge or estimation of the parameters of the acquisition device, nor accurate control of the camera's optics as in SFD, SFF and light-efficient photography (Favaro, 2010; Hasinoff and Kutulakos, 2011; Muhammad and Choi, 2012).

2. User interaction (e.g., the photomontage step in light-efficient photography (Hasinoff and Kutulakos, 2011)).

3. Special hardware, as in coded aperture and plenoptic cameras (Levin et al., 2007; Ng et al., 2005).

The proposed methodology is compatible with any camera with autofocus mechanism (not limited to contrast-based autofocus), as long as a video sequence can be captured during the focusing process.

6.4.2 Algorithm's parameters

The algorithm's parameter, that is, the initial ROI size, has been determined experimentally. The selection of the ROI size is a tradeoff between accuracy in the detection of the focus level and spatial resolution. A large ROI increases the SNR, thus making the estimation of the focus level more reliable. On the other hand, a small ROI will reveal small features in the scene. For all the results shown in this chapter, the ROI size has been set to $1/20th$ of the total frame size. For instance, for images of 640×480 pixels, a ROI of 32×24 pixels have been used.

As for the segmentation algorithm used in the post-processing stage, the parameter of the statistical region merging approach has been fixed to $Q = 64$ for all the experiments. For a detailed description of this parameter and its meaning, the reader is referred to the corresponding references.

6.5 Summary

A new focus-based depth cue has been introduced. The focus measures obtained during the focusing process of a camera, namely the *focus signals*, have been used in order to identify objects at different distances from the camera. The proposed method works by clustering the focus signals without any knowledge of the parameters of the lens during the focusing process. A practical application of the proposed algorithm has been developed by processing autofocus video sequences for obtaining a coarse depth-map of the imaged scene. An extensive set of experiments using different cameras and complex scenes show the potential of the proposed approach.

Due to limitations in the depth-of-field, autofocus is currently an important feature of most off-the-shelf digital cameras. The results in this work show experimental evidence that the autofocus sequences can be successfully exploited in order to retrieve a coarse depth-map of the scene. Although the proposed approach does not provide information about the absolute depth of objects in the scene, the obtained results are potentially useful for scene understanding tasks, such as object segmentation and recognition and depth ordering (Feldman and Weinshall, 2008; Palou and Salembier, 2013), as well as computational photography applications, such as digital refocusing and defocusing (Bae and Durand, 2007; Bando and Nishita, 2007).

In computer vision, the integration of different perceptual cues is an increasingly important research field. In this scope, the proposed approach provides a new framework for the integration of classical perceptual cues, such as texture and color, with the focus cue.

CHAPTER 7

Conclusions

"Any intelligent fool can make things bigger and more complex. It takes a touch of genius — and a lot of courage — to move in the opposite direction."
- Albert Einstein

In this thesis, the focus cue in conventional cameras has been analyzed from both a theoretical and practical perspective. From the theoretical standpoint, the concepts introduced and discussed are of concern for the computer vision community due to the increasingly widespread use of conventional cameras. In general, an accurate understanding and control of the focus in conventional cameras plays a key role in the acquisition and processing of images captured with different devices, ranging from simple cellphone cameras to professional DSLR photographic cameras. From the practical perspective, the concepts introduced in this thesis have been exploited for increasing the speed of autofocusing in contrast-based autofocus, for improved depth estimation through shape-from-focus, for image enhancement through noise-robust focus stacking and for the generation of coarse depth-maps through the new shape-from-autofocus framework. This final chapter presents a summary of the contributions and final remarks of this thesis and suggests future research directions.

7.1 Summary of contributions

The focus of an image acquisition system is of fundamental importance in order to guarantee the fast and effective acquisition of high quality images. Beyond the task of image capture, the focus cue has been exploited in computer vision for the inference of scene features such as depth and shape, as well as for image enhancement tasks through image processing, such as the generation of all-in-focus images.

This thesis has analyzed the focus in digital conventional cameras with motorized lenses. Thus, the results and concepts presented in this dissertation can be applied to a wide range of image acquisition devices, such as cellphone cameras, webcams, compact digital photographic cameras, digital single lens reflex (DSLR) cameras, surveillance cameras, and the like. Specifically, the focus-related applications developed in this thesis include efficient image capture (autofocus and focus sampling), depth estimation (shape-from-focus and shape-from-autofocus) and image enhancement through focus stacking. The main contributions of this thesis are summarized as follows.

7.1.1 Compensation of image magnification shift

In chapter 2, an image processing-based approach for the compensation of the *image magnification shift* problem has been proposed. Depending on the quality of the optics, the magnification shift can have a severe impact on the acquisition of images when changing the focus of the camera due to a side-effect variation of the magnification of the system. Unlike previous approaches, the proposed shift-compensation technique does not require previous calibration nor the storage of feature shift maps, thus being more flexible, practical and easily adaptable to different cameras.

7.1.2 Analysis of focus measure operators

In chapter 3, the factors influencing the performance of *focus measure* operators (i.e., the algorithms used to estimate the degree of focus of an image pixel) have extensively been assessed. An exhaustive set of operators based on different concepts such as the image gradient, image Laplacian, image statistics, discrete cosine transform, and discrete wavelets, among others, have been considered. The obtained results provide experimental support for the conclusion that operators based on similar concepts respond similarly to different imaging factors, such as image noise and the size of the support window (or region of interest). In particular, wavelet-based operators show an improved relative performance for reduced

support windows; and image statistics-based operators are the most robust to increments in the image noise, whereas Laplacian-based and wavelet-based operators are the most sensitive to this factor. These results are of interest for the research community for either the development of new focus measure operators or their practical application.

7.1.3 Efficient focus calibration

Theoretically, the focus of a camera can be understood from two different viewpoints. On the one hand, wave optics provides a thorough description of different phenomena that are fundamental for modeling and explaining the formation of images in a defocused system. On the other hand, the *thin-lens* model, based on a paraxial geometrical optical approximation, has widely been applied in optics and computer vision mostly due to its simplicity and applicability. Notwithstanding, chapters 2 and 4 point out some limitations of the thin-lens model, such as the ambiguity in the near and far limits of the depth-of-field, its dependence on known parameters of the camera and its inefficacy to describe the behavior of a defocused system near perfect focus.

In order to tackle the aforementioned limitations, the classical thin-lens model has been extended in chapter 4 by considering both the diffraction effects, borrowed from wave optics, and the effects of sampling involved in the formation of digital images. This is a fundamental step for proposing new solutions to different focus-related problems. Thus, based on an integral analysis of the image formation process, a new efficient focus calibration methodology has been introduced in chapter 4. In contrast to previous approaches that require complex calibration settings for each focus setting of the camera, the calibration procedure developed in this thesis allows for the simple, efficient and robust calibration along the whole focusing range in a single experiment. This approach has successfully been exploited for improving the speed of autofocusing without the need for explicit knowledge of physical camera parameters.

7.1.4 New focus profile model

As an alternative to the classic concepts of the near and far limits of the depth-of-field, a new *focus profile* model has been introduced. From a theoretical perspective, the focus profile defines a closed-form relationship between the focus level and the parameters of the acquisition device, such as the lens focal length, lens aperture, focus of the camera and target position. This is a valuable information for an effective assessment of the effects of the acquisition parameters on the quality of the captured image. In a practical application, the introduced focus profile

model has been exploited for improving the accuracy of depth estimation through shape-from-focus.

7.1.5 Reliability measure and depth-map carving

The focus profile model proposed in this thesis has further been exploited in chapter 5 through a new method for measuring the reliability of the estimation of the focus level. The proposed *reliability measure* (R-measure) quantifies the confidence of the focus level estimation obtained by means of a focus measure operator. In depth estimation through shape-from-focus, the R-measure has been utilized for successfully discarding inaccurate points in the obtained depth-map (depth-map carving). In contrast to previous approaches, the proposed R-measure does not require the pre-computation of the scene's depth-map or the all-in-focus image and efficiently integrates within the shape-from-focus framework, thus yielding a reduced computational complexity.

7.1.6 Noise-robust focus stacking

In order to tackle the problem of generating all-in-focus images through the fusion of several images with limited depth-of-field, a noise-robust focus stacking algorithm has been presented in chapter 5. In an extensive set of experiments, the proposed *selective all-in-focus* (SAF) algorithm outperformed state-of-the-art alternatives (both from the industry and the research community). The main difference of the SAF algorithm with respect to previous approaches is that it exploits the focus profile in order to generate the pixels of the all-in-focus image instead of spatially filtering the frames of the focus sequence.

7.1.7 Shape from autofocus

An important remark about the focus profile model proposed in this thesis, is that, for a specific set of camera settings, it establishes a one-to-one correspondence between the focus level and the scene geometry (in particular, the pixel depth). This provides a novel interpretation of the autofocusing process of a camera, namely the *shape-from-autofocus* (SFA) approach. SFA yields a coarse estimation of the scene geometry by processing the frames corresponding to an autofocus sequence. In contrast to previous focus-based depth estimation techniques, SFA does not require previous knowledge or calibration of the parameters of the acquisition device nor the focus position of each frame. Although the obtained coarse approximation does not provide absolute depth information, that reconstruction is suitable for computational photography applications, such as digital defocusing, as well as different scene understanding tasks, such as object recognition, object segmentation and

depth ordering. As a final remark, although new models for the description of real physical phenomena require extensive validation under different applications, the models introduced in this thesis accurately describe the observed behavior of real systems. Moreover, their application for solving specific problems have produced promising results. Notwithstanding, a thorough theoretical analysis incorporating concepts from wave optics, as well as from lens design and manufacture, could provide additional insights on the applicability and limitations of the introduced models to a wider set of real problems and capturing devices.

7.2 Future research directions

The concepts and the results presented in this dissertation pave the way for new applications and solutions to different focus-related problems. Some future research directions are summarized below.

7.2.1 Improved focus measure

The measurement of the focus level by means of image processing algorithms is of fundamental interest in computer vision. The errors generated during the estimation of focus can yield wrong results in different tasks, such as automated image acquisition, depth estimation or image enhacenment, among others. In this scope, the analysis performed in chapter 3 regarding focus measure operators and the factors that affect their performance can be extended in several ways. Firstly, by taking advantage of the differential behavior exhibited by focus measure operators according to their working principles, novel focus measure operators can be designed by combining the response of several focus measure operators. In this direction, previous efforts for the efficient integration of texture descriptors by means of different machine learning approaches can be exploited (Melendez, 2010). Secondly, in addition to the image-related variables studied in chapter 3, it would be interesting to identify what texture features are relevant for focus detection. The aim would be to study the response of focus measure operators to different families of microtextures, such as the ones previously proposed by Rao and Lohse (1996).

7.2.2 Closed-form shape-from-defocus

Based on the concepts developed in this thesis, and more precisely on the theoretical focus profile derived in chapter 4, new research efforts are being devoted in order to derive closed-form spatial-domain solutions for the shape-from-defocus

problem. This solution would not only represent a new approach for the shape-from-defocus problem, but could also be integrated into the shape-from-autofocus framework introduced in chapter 6 in order to allow for absolute depth estimation in the generated coarse depth-maps.

7.2.3 Estimation of physical parameters

In chapter 4, the proposed calibration methodology allows the implicit calibration of the focus of a camera as a function of a single parameter, the camera constant κ, thus avoiding the need for the estimating individual parameters (such as the lens focal length, aperture and effective pixel size). In optics, different approaches have been proposed for measuring real physical parameters of the camera, such as the focal length of single lenses (de Angelis et al., 1999; Tay et al., 2005) and compound lens systems (Lei and Dang, 1994; Pahk et al., 2000). However, these methods require experimental settings that prevent them from being applied to conventional cameras. To our knowledge, methods for the experimental calibration of the focus of a camera or the estimation of its different parameters are scarce. In this scope, based on the calibration methodology introduced in chapter 4, a new method for estimating the physical focal length is currently under development.

7.2.4 New calibration methods

The independence of the camera constant κ with respect to both the focus of the camera and the target position u_x opens the possibility for robustly solving the shape-from-focus problem by means of global optimization approaches in order to simultaneously estimate both the depth-map (the target position corresponding to each image pixel) and the camera constant. In this case, the focus calibration problem can be implicitly solved allowing the simultaneous generation of depth-maps as well as the auto-calibration of the focus of the camera.

7.2.5 Improved depth estimation

The concept of reliability introduced in chapter 5 can be interpreted as an effort for tackling the problem of estimating the quality of the obtained results. This is valid when performing either depth estimation or image fusion (through focus stacking). Instead of applying "blind" smoothing or regularization techniques to the obtained results, as in previous approaches, the R-measure aims at assigning a confidence value to the estimated focus measure in order to guide subsequent stages. In this direction, the proposed R-measure can be utilized with the methodology proposed by Liu et al. (2011), which aims at exploiting confidence measures for improving the quality of depth-maps.

7.3 Publications

The following publications have been derived from this thesis:

1. The phase correlation-based approach for the compensation of image magnification shift presented in chapter 1 was **published** in the *International Conference on Pattern Recognition* in August 2010 (Pertuz et al., 2010).

2. The analysis of focus measure operators presented in chapter 3 has been **published** in the *Pattern Recognition* journal (Pertuz et al., 2013e).

3. A paper based on the focus calibration methodology presented in chapter 4 has been **submitted** to the *IEEE Transactions on Image Processing* journal (Pertuz et al., 2013a).

4. A paper based on the focus profile model presented in chapter 4 has been **submitted** to the *IEEE Transactions on Pattern Analysis and Machine Intelligence* journal (Pertuz et al., 2013b).

5. A paper based on the reliability measure (R-measure) and the depth-map carving approach for shape-from-focus presented in chapter 5 has been **submitted** to the *Computer Vision and Image Understanding* journal.

6. A paper based on the selective all-in-focus algorithm presented in chapter 5 has been **published** in the *IEEE Transactions on Image Processing* journal (Pertuz et al., 2013d).

7. A paper based on the shape-from-autofocus algorithm presented in chapter 6 has been **submitted** to the *Computer Vision and Image Understanding* journal (Pertuz et al., 2013c).

APPENDIX A

Focus measure operators

This appendix summarizes the focus measure operators studied in chapter 3 of this dissertation. For homogeneity, the original notation used by some authors has been modified. In addition, the notation has been adapted to explicitly deal with discrete image coordinates. A MATLAB implementation of the studied focus measure operators is pubicly available at http://www.sayonics.com/sources/sff_demo.zip.

A.1 Absolute central moment (MIS1)

Shirvaikar (2004) proposed a focus measure for AF, the *absolute central moment* ($ACMo$), based on statistical measures and the image histogram H:

$$ACMo = \sum_{k=1}^{L} |k - \mu| P_k, \tag{A.1}$$

where μ is the mean intensity value of H, L the number of gray-levels in the image and P_k the relative frequency of the k-th gray-level. This operator has been adapted to SFF by accumulating the values of $ACMo$ computed over the neighborhood $\Omega(x, y)$ of pixel $I(x, y)$.

A.2 Brenner's focus measure (MIS2)

A focus measure based on the second difference of the image gray-levels of an image I is defined as (Firestone et al., 1991; Santos et al., 1997; Sun et al., 2004):

$$\varphi = \sum_{(i,j)} |I(i, j) - I(i + 2, j)|^2 \tag{A.2}$$

A variation in (A.2) also allows taking into account the vertical variations of the image (Santos et al., 1997). In addition, the values above a given threshold can only be accumulated (Santos et al., 1997; Sun et al., 2004). This measure can be adapted to SFF if the focus measure for every pixel $I(x, y)$ is computed by limiting the sum in (A.2) to its local neighborhood $\Omega(x, y)$.

A.3 Image Contrast (MIS3)

Nanda and Cutler (2001) used the image contrast as a focus measure for autofocus:

$$C(x, y) = \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} |I(x, y) - I(i, j)|, \tag{A.3}$$

where $C(x, y)$ is the image contrast for pixel $I(x, y)$. This operator can be adapted to SFF if the contrast is accumulated over the pixel's neighborhood:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} C(i,j), \tag{A.4}$$

A.4 Image curvature measure (MIS4)

This operator was proposed by Helmli and Scherer (2001) for SFF applied to microscopy. If the image gray-levels are interpolated by means of a surface, the curvature of this surface may be used as a focus measure (Helmli and Scherer, 2001; Minhas et al., 2009):

$$\varphi = |c_0| + |c_1| + |c_2| + |c_3|, \tag{A.5}$$

where $C = (c_0, c_1, c_2, c_3)^T$ is the vector of coefficients used to interpolate a quadratic surface $f(x,y) = c_0x + c_1y + c_2x^2 + c_3y^2$. C is computed through least squares by applying two convolution masks (Helmli and Scherer, 2001):

$$c_0 = M_1 * I \qquad c_2 = \frac{3}{2}M_2 * I - M_2^T * I$$
$$c_1 = M_1^T * I \qquad c_3 = \frac{3}{2}M_2^T * I - M_2 * I,$$

where:

$$M_1 = \frac{1}{6}\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad M_2 = \frac{1}{5}\begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

A.5 Helmli and Scherer's mean method (MIS5)

Helmli and Scherer (2001) proposed to measure the local contrast by computing the ratio, $R(x,y)$, between the intensity level of every pixel $I(x,y)$ and the mean gray level of its neighborhood $\mu(x,y)$:

$$R(x,y) = \begin{cases} \dfrac{\mu(x,y)}{I(x,y)}, & \mu(x,y) \geq I(x,y) \\ \dfrac{I(x,y)}{\mu(x,y)}, & \text{otherwise.} \end{cases} \tag{A.6}$$

This ratio is one if there is either a constant gray value or low contrast. An $M \times N$ neighborhood centered at (x,y) is used to compute $\mu(x,y)$. The focus measure for $I(x,y)$ is computed by summing the values of $R(x,y)$ within $\Omega(x,y)$.

A.6 Local Binary Patterns-based measure (MIS6)

Lorenzo et al. (2008) studied the use of Local Binary Patterns (LBP) as a focus measure for autofocus applications. In order to compute the LBP operator for a given pixel $I(x,y)$, n pixels within a radius R around (x,y) are selected (Lorenzo et al., 2008):

$$LBP_{x,y}(n,R) = \sum_{k=1}^{n} S(I_k - I(x,y)), \tag{A.7}$$

where I_k is the intensity level of the k-th pixel around (x,y) and $S(x)$ is:

$$S(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{A.8}$$

The focus measure for pixel $I(x,y)$ is computed as:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} LBP_{i,j}(n,R) \tag{A.9}$$

Values of $n = 8$ and $R = 2$ have been used in the experiments of this dissertation.

A.7 Steerable filters-based measure(MIS7)

Minhas et al. (2009) proposed a focus measure based on a filtered version of the image I_f:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} I_f(i,j), \tag{A.10}$$

$I_f(i,j)$ is defined as:

$$I_f(i,j) = max\{R^{\theta_1}_{(i,j)}, R^{\theta_2}_{(i,j)}, ...R^{\theta_N}_{(i,j)}\}, \tag{A.11}$$

where R^{θ_n}, $n = 1, 2, ...N$, is the image response to the n-th steerable filter defined as (Freeman and Adelson, 1991):

$$R^{\theta_n} = \cos(\theta_n)(I * \Gamma_x) + \sin(\theta_n)(I * \Gamma_y), \tag{A.12}$$

with Γ_x and Γ_y being the Gaussian derivatives (see section A.10).

Recently, an efficient algorithm for the computation of the focus measure based on steerable filters by means of integral images has been proposed by Minhas et al. (2012).

A.8 Spatial frequency measure (MIS8)

This operator was proposed by Huang and Jing (2007) for the fusion of multi-focal images:

$$\varphi_{x,y} = \sqrt{\sum_{(i,j)\in\Omega(x,y)} I_x(i,j)^2 + \sum_{(i,j)\in\Omega(x,y)} I_y(i,j)^2}, \tag{A.13}$$

where I_x and I_y denote the first derivatives of an image in the X and Y direction, respectively.

A.9 Vollath's autocorrelation (MIS9)

A focus measure based on image autocorrelation has been used by Hilsenstein (2005); Santos et al. (1997) and Sun et al. (2004) for autofocus. Its adaptation to SFF is straightforward:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (I(i,j)\cdot I(i+1,j) - \sum_{(i,j)\in\Omega(x,y)} I(i,j)\cdot I(i+2,j) \quad \text{(A.14)}$$

A.10 Gaussian derivative (GRA1)

Based on the defocus modeling, Geusebroek et al. (2000) proposed a focus measure for autofocus in microscopy based on the first order Gaussian derivative (Geusebroek et al., 2000; Russell and Douglas, 2007):

$$\varphi = \sum_{(x,y)} (I * \Gamma_x)^2 + (I * \Gamma_y)^2, \quad \text{(A.15)}$$

where Γ_x and Γ_y are the x and y partial derivatives of the Gaussian function $\Gamma(x,y,\sigma)$, respectively:

$$\Gamma(x,y,\sigma) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \quad \text{(A.16)}$$

In order to apply this measure to small neighborhoods in SFF, the value of σ in (A.16) must be computed accordingly. For the results shown in this dissertation, the value of σ was selected such that, for a neighborhood of size $W\times W$, a total of five σ's are contained along W. The focus measure for a pixel $I(x,y)$ is computed by applying (A.15) within its neighborhood, $\Omega(x,y)$.

A.11 Gradient energy (GRA2)

The sum of squares of the first derivative in the x and y directions has also been proposed as a focus measure (Huang and Jing, 2007; Malik and Choi, 2008; Subbarao et al., 1993):

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (I_x(i,j)^2 + I_y(i,j)^2). \quad \text{(A.17)}$$

A.12 Thresholded absolute gradient (GRA3)

The first derivative of the image in the horizontal dimension is a simple measure of its degree of focus:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} |I_x(i,j)| \quad , |I_x(i,j)| \geq T \quad \text{(A.18)}$$

The performance of this measure is affected by the selection of T. For the sake of generality, no threshold has been considered in this work. An alternative definition of this method considers both vertical and horizontal image derivatives by either addition (Chern et al., 2001) or selection of the maximum value (Santos et al., 1997).

A.13 Squared gradient (GRA4)

Instead of applying (A.18), the first derivative is squared in order to increase the influence of larger gradients (Eskicioglu and Fisher, 1995; Huang and Jing, 2007; Santos et al., 1997; Sun et al., 2004). If both vertical and horizontal derivatives are considered and added, this measure is equivalent to the energy of the image gradient (GRAE).

A.14 3D Gradient (GRA5)

Ahmad and Choi (2007) proposed the use of the 3D gradient as a focus measure operator. In that work, the whole image sequence is stacked in a single image volume $V(x,y,z)$, where x and y denote the image coordinates and z the image number. The magnitude of the 3D gradient is given by:

$$|\nabla V| = \sqrt{\nabla V_x^2 + \nabla V_y^2 + \nabla V_z^2}, \quad \text{(A.19)}$$

where the three components of the gradient are obtained by convolving V with the $3\times3\times3$ operator oriented in the x, y and z direction, respectively. The focus measure at pixel $I(i,j)$ for the k-th image is computed as the sum of the 3D gradient in a small 2D neighborhood, provided this gradient is greater than a threshold T:

$$\varphi_{x,y,k} = \sum_{(i,j)\in\Omega(x,y)} |\nabla V(i,j,k)|, \quad |\nabla V(i,j,k)| \geq T. \quad \text{(A.20)}$$

A.15 Tenengrad (GRA6)

A popular focus measure based on the magnitude of image gradient is defined as (Chern et al., 2001; Helmli and Scherer, 2001; Huang and Jing, 2007; Krotkov and Martin, 1986; Lee et al., 2009, 1995; Malik and Choi, 2008; Minhas et al., 2009; Nair and Stewart, 1992; Pech-Pacheco et al., 2000; Santos et al., 1997; Shen and Chen, 2006; Subbarao et al., 1993; Sun et al., 2004; Wee and Paramesran, 2007; Yang and Nelson, 2003; Yap and Raveendran, 2004):

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} \left(G_x(i,j)^2 + G_y(i,j)^2\right), \quad \text{(A.21)}$$

where G_x and G_y are the X and Y image gradients computed by convolving the given image I with the Sobel operators.

A.16 Tenengrad variance (GRA7)

This operator uses the variance of the image gradient as a focus measure. It was originally used for autofocus by Pech-Pacheco et al. (2000), but can also be applied to SFF:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (G(i,j) - \overline{G})^2, \tag{A.22}$$

where $\overline{G}$ is the mean value within $\Omega(x,y)$ of the gradient magnitude, which in turn is computed as: $G = \sqrt{G_x^2 + G_y^2}$.

A.17 Energy of Laplacian (LAP1)

The energy of the second derivative of the image has been used as a focus measure for both autofocus (Chern et al., 2001; Huang and Jing, 2007; Lee et al., 2009, 1995; Russell and Douglas, 2007; Shen and Chen, 2006; Subbarao et al., 1993; Sun et al., 2004; Wee and Paramesran, 2007; Xie et al., 2006; Yap and Raveendran, 2004) and SFF (Ahmad and Choi, 2007):

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} \Delta I(i,j)^2, \tag{A.23}$$

where ΔI is the image Laplacian obtained by convolving I with the Laplacian mask:

A.18 Modified Laplacian (LAP2)

Nayar and Nakagawa (1994) proposed a focus measure based on an alternative definition of the Laplacian:

$$\varphi(x,y) = \sum_{(i,j)\in\Omega(x,y)} \Delta_m I(i,j), \tag{A.24}$$

where $\Delta_m I$ is the modified Laplacian of I, computed as:

$$\Delta_m I = |I * \mathcal{L}_X| + |I * \mathcal{L}_Y|. \tag{A.25}$$

The convolution masks used to compute the modified Laplacian are:

$$\mathcal{L}_X = \begin{bmatrix} -1 & 2 & -1 \end{bmatrix}$$

and $\mathcal{L}_Y = \mathcal{L}_X^T$.

A.19 Diagonal Laplacian (LAP3)

Thelen et al. (2009) also included vertical variations of the image in order to compute the modified Laplacian of the image:

$$\Delta_m I = |I*\mathcal{L}_X| + |I*\mathcal{L}_Y| + |I*\mathcal{L}_{X1}| + |I*\mathcal{L}_{X2}|, \tag{A.26}$$

where $\mathcal{L}_X$ and $\mathcal{L}_Y$ are defined as in (A.25), and $\mathcal{L}_{X1}$ and $\mathcal{L}_{X2}$ are given by:

$$\mathcal{L}_{X1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathcal{L}_{X2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A.20 Variance of Laplacian (LAP4)

This measure utilizes the variance of the image Laplacian as a focus measure for autofocus (Pech-Pacheco et al., 2000). In SFF, this measure can be defined as:

$$\varphi_{i,j} = \sum_{(i,j)\in\Omega(x,y)} (\Delta I(i,j) - \overline{\Delta I})^2, \tag{A.27}$$

where $\overline{\Delta I}$ is the mean value of the image Laplacian within $\Omega(x,y)$.

A.21 Laplacian in 3D Window (LAP5)

An et al. (2008) proposed the use of a 3D neighborhood for accumulating the focus measure:

$$\varphi_{x,y,k} = \sum_{k=n-1}^{n+1} \sum_{(i,j)\in\Omega(x,y)} |\Delta_M I_k(i,j)|, \tag{A.28}$$

where $\Delta_M I_k$ is the modified Laplacian of the k-th image, computed as in (A.25).

A.22 Chebyshev moments-based (STA1)

A focus measure operator for AF based on Chebyshev moments was proposed by Yap and Raveendran (2004) as the ratio between the energy of the high-pass band and the energy of the low-pass band extracted from the image by using the Chebyshev moments. In (Yap and Raveendran, 2004), this measure is applied to a normalized image $\tilde{I}$ and can be computed as:

$$\varphi = \frac{||\mathcal{H}(\tilde{I};p)||}{||\mathcal{L}(\tilde{I};p)||}, \tag{A.29}$$

where $||\mathcal{H}(\tilde{I};p)||$ and $||\mathcal{L}(\tilde{I};p)||$ respectively denote the high-order and low-order Chebyshev moments up to order p of the normalized image $\tilde{I}$, which is computed as:

$$\tilde{I} = \frac{I}{\sqrt{\sum_{(i,j)} [I(i,j)]^2}} \tag{A.30}$$

Note that (A.29) and (A.30) must be applied to the whole image in order to compute a single focus measure. Nevertheless, this measure can be used in SFF by performing a sliding-block operation within a

neighborhood $\Omega(x, y)$ and assigning the obtained measure to its central pixel. However, this procedure is expected to affect the performance of the operator for small neighborhoods, since the kernels used to compute the Chebyshev moments will lose their discriminating capability as the number of points (window size) is decreased. Parameter p also determines the sensitivity to the frequency components of the image. According to Yap and Raveendran (2004) and Wee and Paramesran (2008a), a value of $p = 2$ has been used in this work.

A.23 Eigenvalues-based (STA2)

A sharpness measure of an image proposed by Wee and Paramesran (2007) is obtained from the trace of the matrix of eigenvalues, Λ, of the image covariance S. Thus, the variances of the principal components of the image are used as a focus measure (Wee and Paramesran, 2007, 2008b,a):

$$\varphi = \text{trace}[\Lambda_k], \tag{A.31}$$

where the trace of Λ_k is the sum of the first k diagonal elements of Λ. k has been set to 5 in this work for neighborhoods equal or greater than 5×5 pixels.

The image covariance S is :

$$S = \frac{JJ^T}{MN - 1}, \tag{A.32}$$

where J is the normalized image in (A.30) after removing its mean value: $J = \tilde{I} - mean(\tilde{I})$; and $M \times N$ is the size of the neighborhood. pixel's neighborhood. This focus measure, originally proposed for a whole image, can be applied to SFF in a sliding block-like fashion. However, the computational cost is dramatically increased since the normalization procedure in (A.30) is iterated for every pixel's neighborhood $\Omega(x, y)$.

A.24 Gray-level variance (STA3)

The variance of image gray-levels is one of the most popular methods to compute the focus measure of an image. It has been applied to both autofocus (Baina and Dublet, 1995; Chern et al., 2001; Firestone et al., 1991; Huang and Jing, 2007; Krotkov and Martin, 1986; Lee et al., 1995; Santos et al., 1997; Shen and Chen, 2006; Subbarao et al., 1993; Sun et al., 2004; Wee and Paramesran, 2007; Xie et al., 2006; Yang and Nelson, 2003; Yap and Raveendran, 2004) and SFF (An et al., 2008; Helmli and Scherer, 2001; Malik and Choi, 2008; Minhas et al., 2009):

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (I(i,j) - \mu)^2, \tag{A.33}$$

where μ is the mean gray-level of pixels within $\Omega(x, y)$.

A.25 Gray-level local variance (STA4)

Pech-Pacheco et al. (2000) proposed the local variance of gray-levels as a focus measure for autofocus of diatoms in brigthfield microscopy. For its application to SFF, this operator is re-formulated as:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (L_v(i,j) - \overline{L_v})^2, \tag{A.34}$$

where $L_v(i, j)$ is computed as the variance of gray-levels within a neighborhood of size $w_x \times w_y$ centered at (i, j). $\overline{Lv}$ is the mean value of L_v. In this work, w_x and w_y have been chosen to coincide with the size of $\Omega(x, y)$.

A.26 Normalized gray-level variance (STA5)

The gray-level variance can be compensated for differences in the average image brightness among different images by normalizing the value of φ in (A.33) by the mean gray-level value μ (Lee et al., 2009; Santos et al., 1997; Sun et al., 2004).

A.27 Modified gray-level variance (STA6)

The computation of the gray-level variance in (A.33) can be thought of as a non-linear filtering of the image. An alternative focus measure can be obtained if the mean value $\mu(x, y)$ of every pixel within its neighborhood $\Omega(x, y)$ is computed:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (I(i,j) - \mu(i,j))^2, \tag{A.35}$$

where $\mu(x, y)$ is obtained through an averaging filter.

A.28 Histogram entropy (STA7)

Since a focused image is expected to have a higher information content, the entropy and range of the image histogram can be used to compute the focus measure. The histogram entropy operator is defined as (Chern et al., 2001; Firestone et al., 1991; Krotkov and Martin, 1986; Santos et al., 1997; Sun et al., 2004; Xie et al., 2006):

$$\varphi = -\sum_{k=1}^{L} P_k \log(P_k), \tag{A.36}$$

where P_k is the relative frequency of the k-th gray-level.

In order to compute a focus value for a pixel at coordinates (x, y), the image histogram used in (A.36) is obtained from the gray-level values within $\Omega(x, y)$.

A.29 Histogram Range (STA8)

The histogram range has been used as a focus measure for autofocus (Firestone et al., 1991; Santos et al., 1997; Sun et al., 2004):

$$\varphi = max(k|H > 0) - min(k|H > 0) \tag{A.37}$$

In this work, the histogram H is computed within every $\Omega(x, y)$.

A.30 DCT energy ratio (DCT1)

The discrete cosine transform (DCT) is now part of many image and video encoding systems. As noted by Baina and Dublet (1995), the sum of the AC components of the DCT is equal to the variance of the image intensity and can be used as a focus measure. Later, Shen and Chen (2006) proposed the DC/AC ratio as a focus measure. Let $F_{u,v}$ be the DCT of an $M \times N$ sub-block of the image (typically, M = N = 8). The focus measure associated with this sub-block, φ_S, can be computed as:

$$\varphi_S = \frac{\sum_{u=0}^{M-1}\sum_{\substack{v=0 \\ (u,v)\neq(0,0)}}^{N-1} F(u,v)^2}{F_{0,0}^2}, \tag{A.38}$$

For SFF, the focus measure for a pixel $I(x, y)$ is computed by accumulating the values of φ_S within its neighborhood $\Omega(x, y)$.

A.31 DCT reduced energy ratio (DCT2)

Based on a statistical analysis of the information content of in the DCT coefficients, Lee et al. (2009) applied the DCT to 8×8 image sub-blocks in order to measure focus. They suggested that the computation time and robustness to noise of the energy ratio measure in (A.38) can be improved if only 5 out of the 63 AC coefficients are used to compute the AC energy. Thus, the focus measure is defined as:

$$\varphi = \frac{F_{0,1}^2 + F_{1,0}^2 + F_{2,0}^2 + F_{1,1}^2 + F_{0,2}^2}{F_{0,0}^2}$$

A.32 Modified DCT (DCT3)

An efficient implementation of a focus measure based on an 8×8 modified DCT can be obtained by performing a linear convolution with a mask $\mathcal{M}$ (Lee et al., 2008). Similarly to DCTR and DCTE, the focus measure for SFF is computed for every pixel according to its neighborhood:

$$\varphi_{x,y} = \sum_{(i,j)\in\Omega(x,y)} (I * \mathcal{M}), \tag{A.39}$$

where

$$\mathcal{M} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

A.33 Sum of Wavelet coefficients (WAV1)

Wavelet-based focus measure operators are mostly based on the statistical properties of the discrete wavelet transform (DWT) coefficients. In the first level DWT, the image is decomposed into four sub-images, where W_{LH1}, W_{HL1}, W_{HH1} and W_{LL1} denote the three detail sub-bands and the coarse approximation sub-band, respectively. For a higher level DWT, the coarse approximation is successively decomposed into detail and coarse sub-bands. The information of the detail and coarse sub-bands is then used to compute the focus measure.

Yang and Nelson (2003) proposed a focus operator for autofocus computed from the sub-bands:

$$\varphi = \sum_{(i,j)\in\Omega_D} |W_{LH1}(i,j)| + |W_{HL1}(i,j)| + |W_{HH1}(i,j)|, \tag{A.40}$$

where Ω_D is the corresponding window of Ω in the DWT sub-bands. In this dissertation, the focus measure of all the wavelet-based operators has been computed using the coefficients of the over-complete wavelet transform, thus avoiding the need for computing the corresponding neighborhood within each sub-band. Thus, Ω_D is simply the same as Ω.

Huang et al. (2005) used a focus measure similar to (A.40) with 2-level DWT and Daubechies-10 filters. In this work, 1-level DWT with Daubechies-6 filters have been used by following (Yang and Nelson, 2003) and (Xie et al., 2006).

A.34 Variance of Wavelet coefficients (WAV2)

The variance of the wavelet coefficients within Ω_D can also be used to compute the focus measure (Yang and Nelson, 2003):

$$\begin{aligned} \varphi = & \sum_{(i,j)\in\Omega_D} (W_{LH1}(i,j) - \mu_{LH1})^2 \\ & + \sum_{(i,j)\in\Omega_D} (W_{HL1}(i,j) - \mu_{HL1})^2 \\ & + \sum_{(i,j)\in\Omega_D} (W_{HH1}(i,j) - \mu_{LL1})^2, \end{aligned} \tag{A.41}$$

where μ_{LH}, μ_{HL} and μ_{HH} denote the mean value of the respective DWT sub-bands within Ω_D.

A.35 Ratio of wavelet coefficients (WAV3)

Xie et al. (2006) proposed the use of the ratio between the high frequency coefficients M_H and the low frequency coefficients M_L of the Wavelet transform as a focus measure (Xie et al., 2006):

$$\varphi = \frac{M_H^2}{M_L^2}, \tag{A.42}$$

where M_H and M_L are defined as follows:

$$M_H^2 = \sum_k \sum_{(i,j)\in\Omega_D} W_{LHk}(i,j)^2 + W_{HLk}(i,j)^2 + W_{HHk}(i,j)^2 \tag{A.43}$$

$$M_L^2 = \sum_k \sum_{(i,j)\in\Omega_D} W_{LLk}(i,j)^2. \tag{A.44}$$

Sub-index k indicates that the k-th level wavelet is used to compute the coefficients. According to (Xie et al., 2006), the coefficients of the first level DWT are used in (A.43), whereas the third level coefficients are used in (A.44). The WAV1, WAV2 and WAV3 operators were originally proposed for autofocus applications. In order to adapt them to SFF, a focus measure is computed for every pixel $I(x, y)$ by restricting the sums in (A.40)-(A.42) to the corresponding $\Omega(x, y)$.

A.36 Ratio of curvelet coefficients

Minhas et al. (2011) proposed a focus measure operator based on the coefficients of the discrete curvelet transform. In the k-th level, the curvelet transform decomposes an image into N bands at different orientations. Similarly to the wavelet-based focus measure operators described previously, the focus measure is computed as:

$$\varphi = \sum_{(i,j)\in\Omega_D} F_\theta(i,j), \tag{A.45}$$

where $F_\theta(i,j)$ is calculated as the ratio between the summed coefficients of the k-th and $(k-1)$-th level sub-bands. Let C_k denote the coefficients of the k-th sub-band, $F_\theta(i,j)$ is defined as:

$$F_\theta(i,j) = \frac{\sum C_k(i,j)}{\sum C_{k-1}(i,j)} \tag{A.46}$$

Following (Minhas et al., 2011), 2-level curvelet decomposition with eight orientations has been implemented. In order to perform a fair comparison with other focus measure operators, the pre-processing steps of contrast enhancement and denoising described in (Minhas et al., 2011) have been ommited.

APPENDIX B

Defocus simulation

This appendix presents the defocus simulation algorithm used to generate synthetical focus sequences in this dissertation. The defocus model is based on the paraxial geometrical approximation of defocus using a Gaussian point spread function. In order to overcome the shift-invariant restriction of the isoplanatic assumption, the defocused image is generated by simulating defocus on each point of the source radiance.

B.1 Defocus blur

According to the linear shift-invariant model of focus (chapter 2), a defocused image is computed by filtering source radiance with a blurring kernel known as the point spread function (PSF). Thus, the defocused image I_D can be described as the convolution of the focused one I with a blurring function h:

$$I_D = I * h \tag{B.1}$$

In incoherent polychromatic illumination, the PSF can be simplified as a 2D Gaussian:

$$h(\omega, \nu) = \frac{1}{2\pi\rho_{x,y}^2} \exp\left(-\frac{\omega^2 + \nu^2}{2\rho_{x,y}}\right), \tag{B.2}$$

where $\sigma_{x,y}$ is proportional to the degree of focus and depends on the depth, $z(x, y)$, of the point at coordinates (x, y). In pixels, the blur parameter $\rho_{x,y}$ can be computed as (chapter 4):

$$\rho_{x,y} = \frac{\gamma f^2}{N} \frac{|z(x, y) - u|}{z(x, y)(u - f)}, \tag{B.3}$$

where u is the current focus of the camera, γ is a constant that depends on the real pixel size, f the focal length of the camera and N the f-number.

B.2 Shift-variant defocus

The convolution in (B.1) is only valid under the assumption of a spatially invariant blurring function within the evaluation window (isoplanatism). Therefore, in order to avoid the isoplanatic restriction, the blurred image is composed from the blurred sub-images $B_{x,y}$ corresponding to every scene point. The blurred sub-image $B_{x,y}$ for a point a coordinates (x, y) is computed by convolving it with the corresponding PSF:

$$B_{x,y} = I(x, y) * h(x, y), \tag{B.4}$$

where $h(x, y)$ denotes the PSF corresponding to pixel $I(x, y)$ according to its depth (found by replacing (B.2) in (B.1)). In turn, the defocused image for the pixel at

(x_0, y_0) is obtained by summing ip the contribution of every sub-image:

$$I_D(x_0, y_0) = \int \int_{\forall(x,y)} B_{x,y}(x-x_0, y-y_0)\, dx\, dy \quad \text{(B.5)}$$

In the above equations, however, since every point is linearly convolved with its corresponding PSF, the overall processing is non-linear and allows the definition of a shift-variant PSF at the cost of a high computational load. The computation time can be reduced by taking into account that not all of the blurred images $B_{x,y}$ must be considered in (B.5), since the radiance of every depicted point spreads only over a small image area depending on the value of its corresponding $\sigma_{x,y}$. As a result, the values of $h_{x,y}$ can be neglected for pixel coordinates beyong 2.5 standard deviations away from (x_0, y_0). Thus the integral in (B.5) can be limited to those pixels that comply with:

$$(x - x_0)^2 + (y - y_0)^2 \leq 6.25\rho_{x,y}^2 \quad \text{(B.6)}$$

B.3 Image noise

In order to consider the effect of noise, two noise components are added to the defocused image: a radiance dependent noise $n(I)$ and a radiance-independent component n. Thus, the noisy image I_n is computed from the ideal defocused image I_D as: $I_n = n(I_D) + n$. The radiance-independent component, n, corresponds to a Gaussian noise with zero mean and variance $\sigma_n = \nu$, and the radiance-dependent noise is a Gaussian noise with zero mean and a shift-variant variance $\sigma_{x,y} = \sqrt{\nu} I(x, y)$. The noise parameter, ν, corresponding to the i-th noise level is defined as:

$$\nu = 0.6i \times 10^{-3} + 0.5 \times 10^{-3} \quad \text{(B.7)}$$

APPENDIX C

Focus profile: error sources

This appendix analyzes different error sources on the computation of the focus profile. In particular, the variation of the camera constant along the image field as well as the effects of the error associated with the estimation of the camera constant are discussed. The experiments provided in this appendix suggest that, although the characteristics of an optical system change along the image field - mostly due to optical aberrations – this yields a negligible change on the camera constant.

C.1 Optical aberrations

As stated in chapter 2, optical aberrations can affect the perceived focus level for a given focus setting. As a result, the estimated camera constant may vary along the image field depending on the distance from the optical axis and the amount of distortion of the camera. Arguably, the change in the focus level along the image field should be negligible when compared to its variation as a function of the focus of the camera. This fact is illustrated with the following experiment.

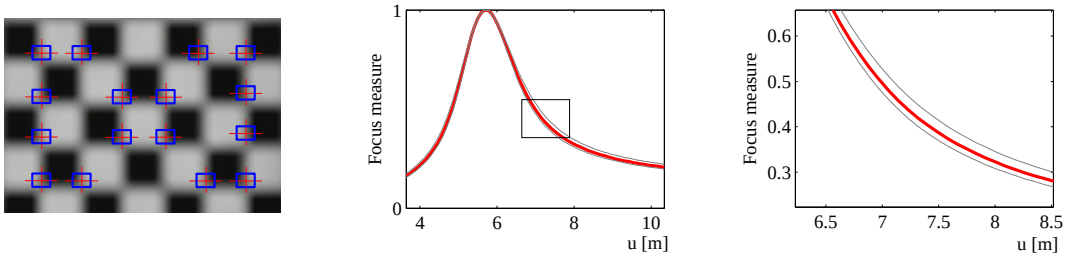
Fig. C.1a shows a checker board pattern used to compute the camera constant at different image positions of the image field according to the calibration procedure presented in chapter 4. In this experiment, the average camera constant along the image field is $\kappa = 109.8$, with maximum and minimum variations of $+6.7\%$ and -4.7%, respectively.

As shown in Fig. C.1b, the variations of the camera constant along the image field yield small variations on the values of the focus profile (gray curves). Specifically, the focus profile has a maximum variation between -3.8% and $+6.01\%$ along the whole focusing range. In the scope of this thesis, this can be neglected for several applications such as focus stacking, shape-from-focus and autofocus. Notwithstanding, in the presence of severe image aberrations that could affect the estimation of the focus level, this effect should be carefully assessed. For the particular case of severe field curvature aberration, its effect could be evidenced by a slight shift of the position of the peak of the focus profile.

C.2 Calibration error

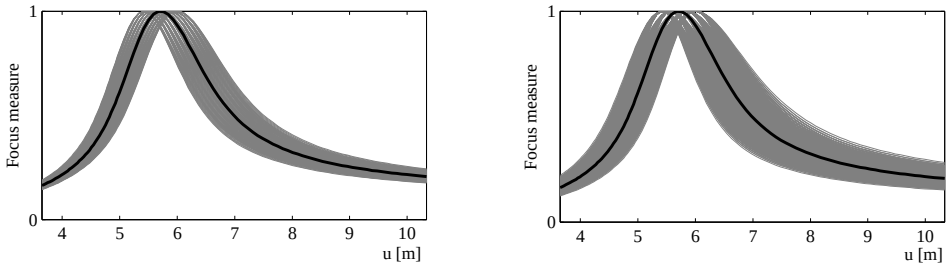
The parameters estimated during the calibration proposed in chapter 4 are the camera constant, κ, and the target position, u_x. The uncertainty on the corresponding focus profile, $\delta\tilde{\varphi}$, is given by (Taylor, 1997):

$$(\delta\tilde{\varphi})^2 = \left(\frac{\partial\varphi}{\partial\kappa}\delta\kappa\right)^2 + \left(\frac{\partial\varphi}{\partial u_x}\delta u_x\right)^2, \qquad \text{(C.1)}$$



(a) (b) (c)

Figure C.1: Optical aberrations. (a) Different poisitions on the image field used for calibration. (b) Extremes of the variation of the focus profile. (c) Zoom on (b).



(a) (b)

Figure C.2: The focus profile uncertainty. (a) Uncertainty of the focus profile for $\kappa \pm 5\%$ and $u_x \pm 5\%$. (b) Uncertainty of the focus profile for $\gamma \pm 5\%$, $N \pm 5\%$, $f \pm 5\%$ and $u_x \pm 5\%$.

where $\delta\kappa$ and δu_x are the uncertainties in the estimation of the camera constant and the target position, respectively.

It can be readily verified from (C.1) that the error in the focus profile is stable with respect to its parameters. This is illustrated in Fig. C.2a, which shows the variations of the focus profile for $\kappa \pm 5\%$ and $u_x \pm 5\%$. Interestingly enough, the estimation of the focus profile as a function of the camera constant halves the uncertainty with respect to its estimation as a function of individual camera parameters (focal length, pixel density and f-number), as shown in Fig. C.2b.

Bibliography

Adelson, E. and Wang, J. (1992). Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106.

Aggarwal, M. and Ahuja, N. (2002). A pupil-centric model of image formation. *International Journal of Computer Vision*, 48(3):195–214.

Aguet, F., Van De Ville, D., and Unser, M. (2008). Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy. *IEEE Transactions on Image Processing*, 17(7):1144–1153.

Ahmad, M. B. and Choi, T. S. (2005). A heuristic approach for finding best focused shape. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):566–574.

Ahmad, M. B. and Choi, T. S. (2007). Application of three dimensional shape from image focus in LCD/TFT displays manufacturing. *IEEE Transactions on Consumer Electronics*, 53(1):1–4.

Allen, E. and Traintaphillidou, S. (2011). *The manual of photography*. Elsevier Ltd., 10th edition.

An, Y., Kang, G., Kim, I. J. J., Chung, H. S., and Park, J. (2008). Shape from focus through laplacian using 3D window. In *proc. International Conference on Future Generation Communication and Networking*, volume 2, pages 46 –50.

Antunes, M., Trachtenberg, M., Thomas, G., and Shoa, T. (2005). All-in-focus imaging using a series of images on different focal planes. In *Image Analysis and Recognition*, volume 3656 of *LNCS*, pages 174–181. Springer.

Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916.

Asif, M. and Choi, T. S. (2001). Shape from focus using multilayer feedforward neural networks. *IEEE Transactions on Image Processing*, 10(11):1670–1675.

Aslantas, V. and Kurban, R. (2010). Extending depth-of-field of a digital camera using particle swarm optimization based image fusion. In *IEEE International Symposium on Consumer Electronics*, page 1–5.

Atkins, J., Fiser, J., and Jacobs, R. (2000). Experience-based visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, 41:449–461.

Aydin, T. and Akgul, Y. (2008). A new adaptive focus measure for shape from focus. In *British Machine Vision Conference*.

Baba, M., Asada, N., and Oda, A. (2001). Depth from blur by zooming. In *proc. Vision Interface Annual Conference*, pages 165–172.

Baba, M., Asada, N., Oda, A., and Migita, T. (2002). A thin lens based camera model for depth estimation from defocus and translation by zooming. In *proc. International Conference on Vision Interface*, pages 274–281.

Baba, M., Oda, A., Asada, N., and Yamashita, H. (2006). Depth from defocus by zooming using thin lens-based zoom model. *Electronics and Communications in Japan, part 2*, 89(9):53–62.

Bae, S. and Durand, F. (2007). Defocus magnification. In *proc. Eurographics 2007*, volume 26.

Baina, J. and Dublet, J. (1995). Automatic focus and iris control for video cameras. In *proc. International Conference on Image Processing and its Application*, pages 232–235.

Bando, Y. and Nishita, T. (2007). Towards digital refocusing from a single photograph. In *15th Pacific Conference on Computer Graphics and Applications*, pages 363–372.

Baradarani, A., Wu, Q. J., Ahmadi, M., and Mendapara, P. (2012). Tunable halfband-pair wavelet filter banks and application to multifocus image fusion. *Pattern Recognition*, 45(2):657 – 671.

Barron, J. T. and Malik, J. (2012). Shape, albedo, and illumination from a single image of an unknown object. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 334–341.

Basokur, A. (1998). Digital filter design using hyperbolic tangent functions. *J. Balkan Geophys. Soc.*, 1:14–18.

Bass, M., editor (2010). *Handbood of Optics, Geometrical and Physical optics, Polarized light, Components and Instruments*, volume 1. OSA, 3rd edition.

Bergholm, F. (1987). Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(6):726 –741.

Berriel, L. R., Bescos, J., and Santisteban, A. (1983). Image restoration for a defocused optical system. *Applied Optics*, 22(18):2772–2780.

Biersdorf, W. R. and Baird, J. C. (1966). Effects of an artificial pupil and accommodation on retinal image size. *Journal of the Optical Society of America*, 56(8):1123–1129.

Bilcu, R., Alenius, S., and Vehvilainen, M. (2009). A novel method for multi-focus image fusion. In *proc. IEEE International Conference on Image Processing*, pages 1525 –1528.

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

Born, M. and Wolf, E. (1999). *Principles of Optics*. Cambridge University Press, 7th edition.

Brenner, K. H., Lohmann, A. W., and Ojeda-Castaneda, J. (1983). The ambiguity function as a polar display of the OTF. *Optics Communications*, 44(5):323 – 326.

Breuss, M., Vogel, O., and Tankus, A. (2011). Modern shape from shading and beyond. In *proc. IEEE International Conference on Image Processing*, pages 1–4.

Brown, D. C. (1971). Close-range camera calibration. *Photogrammetric Engeneering*, 37:855–866.

Buades, A., Coll, B., and Morel, J.-M. (2005a). A non-local algorithm for image denoising. In *proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 60 – 65 vol. 2.

Buades, A., Coll, B., and Morel, J. M. (2005b). A review of image denosing algorithms, with a new one. *Multiscale Modeling & Simulation*, 4:490–530.

Burch, C. R. (1942). On the optical see-saw diagram. *Monthly Notices of the Royal Astronomical Society*, 102:159.

Burge, J. and Geisler, W. S. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences*, 108(40):16849–16854.

Burt, P. and Kolczynski, R. (1993). Enhanced image capture through fusion. In *proc. International Conference on Computer Vision*, pages 173 –182.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.

Cao, G., Zhao, Y., and Ni, R. (2010). Edge-based blur metric for tamper detection. *Journal of Information Hiding and Multimedia Signal Processing*, 1(1):20–27.

Chan, T. and Wong, C.-K. (1998). Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370 –375.

Chang, S., Yu, B., and Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532 –1546.

Chen, S. Y. and Li, Y. F. (2013). Finding optimal focusing distance and edge blur distribution for weakly calibrated 3D vision. *IEEE Transactions on Industrial Informatics*, page (in print).

Chen, Y. S., Shih, S. W., Hung, Y. P., and Fuh, C. S. (2000). Camera calibration with a motorized zoom lens. In *proc. International Conference on Pattern Recognition*, volume 4, pages 495 –498.

Chern, N. N. K., Neow, P. A., and Ang, M. H. (2001). Practical issues in pixel-based autofocusing for machine vision. In *proc. IEEE International Conference on Robotics and Automation*, volume 3, pages 2791–2796.

Cody, W. J. (1969). Rational chebyshev approximations for the error function. *Mathematics of Computation*, 23(107):631–637.

Corle, T. R. and Kino, G. S. (1996). *Confocal Scanning Optical Microscopy and Related Imaging Systems*. Academic Press.

Darrell, T. and Wohn, K. (1988). Pyramid based depth from focus. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 504–509.

de Angelis, M., Nicola, S. D., Ferraro, P., Finizio, A., Pierattini, G., and Hessler, T. (1999). An interferometric method for measuring short focal length refractive lenses and diffractive lenses. *Optics Communications*, 160(1-3):5 – 9.

DeCarlo, D. (2002). Towards real-time cue integration by using partial results. In *proc. European Conference on Computer Vision*.

Denton, M. B., editor (2000). *Further developments in scientific optical imaging*. Royal Society of Chemistry.

Dowski, E. R. and Cathey, W. T. (1994). Single-lens, single-image, incoherent passive ranging systems. *Applied Optics*, 33:6762–6773.

Dowski, E. R. and Cathey, W. T. (1995). Extended depth of field through wave-front coding. *Applied Optics*, 34:1859–1866.

Engelhardt, K. and Knop, K. (1988). Acquisition of 3-d data by focus sensing. *Applied Optics*, 27:4684–4689.

Ens, J. and Lawrence, P. (1993). An investigation of methods for determining depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):97 –108.

Ersoy, O. K. (2007). *Diffraction, Fourier Optics and Imaging*. Wiley & Sons.

Eskicioglu, A. M. and Fisher, P. S. (1995). Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965.

Fagueras, O., Luong, Q. T., and Maybank, S. J. (1992). Camera self-calibration theory and experiments. In *proc. European Conference on Computer Vision*, volume 588, pages 321–334.

Favaro, P. (2007). Shape from focus and defocus: Convexity, quasiconvexity and defocus-invariant textures. In *proc. IEEE International Conference on Computer Vision*, pages 1–7.

Favaro, P. (2010). Recovering thin structures via nonlocal-means regularization with application to depth from defocus. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1133 –1140.

Favaro, P., Burger, M., and Soatto, S. (2004). Scene and motion reconstruction from defocused and motion-blurred images via anisotropic diffusion. In *proc. European Conference on Computer Vision*, pages 257–269.

Favaro, P., Menucci, A., and Soatto, S. (2003). Observing shape from defocused images. *International Journal of Computer Vision*, 52(1):25–43.

Favaro, P. and Soatto, S. (2005). A geometric approach to shape from defocus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):406 –417.

Favaro, P. and Soatto, S. (2006). *3D Shape Estimation and Image Restoration: Exploiting Defocus and Motion Blur*. Springer-Verlag.

Favaro, P., Soatto, S., Burger, M., and Osher, S. (2008). Shape from defocus via difussion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):518–531.

Feldman, D. and Weinshall, D. (2008). Motion segmentation and depth ordering using an occlusion detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1171–1185.

Fenimore, E. E. and Cannon, T. M. (1978). Coded aperture imaging with uniformly redundant arrays. *Applied Optics*, 17:337.

Fergus, R., Singh, B., Hertzmann, A., Roweis, S. T., and Freeman, W. T. (2006). Removing camera shake from a single photograph. *ACM Transactions on Graphics*, 25:787–794.

Fiete, R. (2010). *Modeling the Imaging Chain of Digital Cameras*. SPIE press.

Firestone, L., Cook, K., Culp, K., Talsania, N., and Jr., K. P. (1991). Comparison of autofocus methods for automated microscopy. *Cytometry*, 12(3):195–206.

FitzGerrell, A. R., Edward R. Dowski, j., and Cathey, W. T. (1997). Defocus transfer function for circularly symmetric pupils. *Applied Optics*, 36(23):5796–5804.

Flint, A., Murray, D., and Reid, I. (2011). Manhattan scene understanding using monocular, stereo, and 3D features. In *proc. IEEE International Conference on Computer Vision*, pages 2228 –2235.

Florea, C. and Florea, L. (2011). A parametric non-linear algorithm for contrast based auto-focus. In *IEEE International Conference on Intelligent Computer Communication and Processing*, pages 267 –271.

Foroosh, H., Zenubia, J., and Berthod, M. (2002). Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3):188–200.

Forster, B., Van De Ville, D., Berent, J., Sage, D., and Unser, M. (2004). Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images. *Microscopy Research and Technique*, 65(1-2):33–42.

Fraser, C. S. and Al-Ajlouni, S. (2006). Zoom-dependent camera calibration in digital close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 72(9):1017–1026.

Freeman, W. T. and Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:891–906.

Freund, Y. and Schapire, R. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory*, volume 904 of *LNCS*, pages 23–37. Springer Berlin / Heidelberg.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 28(2):337–407.

Gaganov, V. and Ignatenko, A. (2009). Robust shape from focus via Markov random fields. In *proc. International Conference on Computer Graphics and Vision*, pages 74–80.

Gamadia, M. and Kehtarnavaz, N. (2012). A filter-switching auto-focus framework for consumer camera imaging. *IEEE Transactions on Consumer Electronics*, 58(2):228–236.

Gaskill, J. D. (1978). *Linear Systems, Fourier Transforms, and Optics*. John Willey & Sons.

Geusebroek, J. M., Cornelissen, F., Smeulders, A. W. M., and Geerts, H. (2000). Robust autofocusing in microscopy. *Cytometry*, 39:1–9.

Gibson, J. J. (1950). The perception of visual surfaces. *The American Journal of Psychology*, 63:367–384.

Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall, 3rd edition.

Goodman, J. W. (1996). *Introduction to Fourier optics*. McGraw-Hill, 2nd edition.

Gopinath, R., Odegard, J., and Burrus, C. (1994). Optimal wavelet representation of signals and the wavelet sampling theorem. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 41(4):262 –277.

Groen, F. C. A., Young, I. T., and Ligthart, G. (1985). A comparisong of different focus functions for use in autofocus algorithms. *Cytometry*, 6(2):81–91.

Haralick, R. M. (1984). Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):58 –68.

Hart, J. F. (1968). *Computer approximations*. John Willey & Sons.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition.

Hasinoff, S. (2009). Confocal stereo. *International Journal of Computer Vision*, 81:82–104.

Hasinoff, S. and Kutulakos, K. (2011). Light-efficient photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2203 –2214.

Hasinoff, S. W. (2008). *Variable-aperture photography*. PhD thesis, University of Toronto.

Hasinoff, S. W., Kutulakos, K. N., Durand, F., and Freeman, W. T. (2009). Time-constrained photography. In *proc. IEEE International Conference on Computer Vision*, pages 333 –340.

He, J., Zhou, R., and Hong, Z. (2003). Modified fast climbing search auto-focus algorithm with adaptive step size searching technique for digital camera. *IEEE Transactions on Consumer Electronics*, 49(2):257 – 262.

Healey, G. and Kondepudy, R. (1994). Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276.

Heikkila, J. and Silven, O. (1996). Calibration procedure for short focal length off-the-shelf CCD cameras. In *proc. International Conference on Pattern Recognition*, pages 166–170.

Heikkila, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1106–1112.

Helicon Soft (2011). Helicon focus (v3). retrieved on 01/04/2011 from http://www.heliconsoft.com/heliconfocus.html.

Helmli, F. and Scherer, S. (2001). Adaptive shape from focus with an error estimation in light microscopy. In *proc. International Symposium on Image and Signal Processing and Analysis*, pages 188–193.

Hilsenstein, V. (2005). Robust autofocusing for automated microscopy imaging of fluorescently labelled bacteria. In *proc. Digital Image Computing: Techniques and Applications*, pages 15–15.

Hopkins, H. H. (1955). The frequency response of a defocused optical system. *proc. Royal Society of London*, 231:91–103.

Horn, B. (1968). Focusing. Technical report, MIT.

Horn, B. (1975). *Obtaining shape from shading information*. McGraw-Hill.

Horn, B. K. P. (1990). *Robot Vision*. MIT Press, 6th edition.

Hornberg, A. (2006). *Handbook of Machine Vision*. Wiley-VCH.

Huang, J., Shen, C., Phoong, S., and Chen, H. (2005). Robust measure of image focus in the wavelet domain. In *proc. Int. Symposium on Intelligent Signal Processing and Communication Systems*, pages 157–160.

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

Hwang, T. l., Clark, J. J., and Yuille, A. L. (1989). A depth recovery algorithm using defocus information. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 476 –482.

Jähne, B. (2004). *Practical Handbook on Image Processing for Scientific and Thechnical Applications*. CRC Press, 2nd ed. edition.

Jarvis, R. A. (1983). A perspective on range finding techniques for computer vision. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 5(2):122–139.

Jeon, J., Yoon, I., Kim, D., Lee, J., and Paik, J. (2010). Fully digital auto-focusing system with automatic focusing region selection and point spread function estimation. *IEEE Transactions on Consumer Electronics*, 56(3):1204 –1210.

Ji, H. and Wang, K. (2012). Robust image deblurring with an inaccurate blur kernel. *IEEE Transactions on Image Processing*, 21(4):1624 –1634.

Jia, Z., Gallagher, A., Chang, Y. J., and Chen, T. (2012). A learning-based framework for depth ordering. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 294–301.

Joshi, N., Szeliski, R., and Kriegman, D. (2008). PSF estimation using sharp edge prediction. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*.

Judy, P. F. (1976). The line spread function and modulation transfer function of a computed tomographic scanner. *Medical Physics*, 3:233–236.

Juneja, M. and Sandhu, P. S. (2009). Performance evaluation of edge detection techniques for images in spatial domain. *International Journal of Computer Theory Engeneering*, 1(5):614–621.

Kayargadde, V. and Martens, J.-B. (1996). Estimation of perceived image blur using edge features. *International Journal of Imaging Systems and Technology*, 7(2):102–109.

Kehtarnavaz, N. and Oh, H. J. (2003). Development and real-time implementation of a rule-based auto-focus algorithm. *Real-Time Imaging*, 9:197–203.

Kim, S., Lee, E., Hayes, M. H., and Paik, J. (2012). Multifocusing and depth estimation using a color shift model-based computational camera. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*, 21:4152–4166.

Kim, S. J. and Pollefeys, M. (2008). Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4):562 –576.

Kinba, S. A., Hamada, O. M., Ueda, H. H., Sugitani, Y. K., and Ootsuka, S. H. (Jan. 28, 1997). Auto focus detecting device comprising both phase-difference and contrast detecting methods (patent).

Kodama, K., Mo, H., and Kubota, A. (2006). Free viewpoint, iris and focus image generation by using a three-dimensional filtering based on frequency analysis of blurs. In *proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 625–628.

Kodama, K., Mo, H., and Kubota, A. (2007). Simple and fast all-in-focus image reconstruction based on three-dimensional/two-dimensional transform and filtering. In *proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 769–772.

Krotkov, E. and Martin, J. P. (1986). Range from focus. In *proc. IEEE International Conference on Robotics and Automation*, volume 3, pages 1093 – 1098.

Kuo, C. F. J. and Chiu, C. H. (2011). Improved auto-focus search algorithms for CMOS image-sensing module. *Journal of Information Science and Engineering*, 27:1377–1393.

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

Lai, Y. C. (2011). PSO-based estimation for Gaussian blur in blind image deconvolution problem. In *proc. IEEE International conference on Fuzzy Systems*, pages 1143–1148.

Laikin, M. (2006). *Lens Design*. CRC Press.

Lam, E. Y. and Goodman, J. (2000). A mathematical analysis of the DCT coefficient distributions for images. *IEEE Transactions on Image Processing*, 9(10):1661–1666.

Lee, J., Kim, K., Nam, B., Lee, J., Kwon, Y., and Kim, H. (1995). Implementation of a passive automatic focusing algorithm for digital still camera. *IEEE Transactions on Consumer Electronics*, 41(3):449–454.

Lee, S. Y., Kumar, Y., Cho, J. M., Lee, S. W., and Kim, S. W. (2008). Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning. *IEEE Transactions on Circuits and Systems and Video Technology*, 18(9):1237–1246.

Lee, S. Y., Yoo, J. T., Kumar, Y., and Kim, S. W. (2009). Reduced energy-ratio measure for robust autofocusing in digital camera. *IEEE Signal Processing Letters*, 16(2):133 –136.

Lei, F. and Dang, L. K. (1994). Measuring the focal length of optical systems by grating shearing interferometry. *Applied Optics*, 33(28):6603–6608.

Lenz, M., Ferstl, D., Ruther, M., and Bischof, H. (2012). Depth coded shape from focus. In *proc. International Conference on Computational Photography*.

Levin, A., Fergus, R., Durand, F., and Freeman, W. T. (2007). Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3).

Levoy, M. (2011). Digital photography: lecture notes. Online http://graphics.stanford.edu/courses/cs178-10/applets/autofocusPD.html, Stanford University.

Levoy, M., Ng, R., Adams, A., Footer, M., and Horowitz, M. (2006). Light field microscopy. *ACM Transactions on Graphics*, 25(3):924–934.

Li, M. and Lavest, J. M. (1996). Some aspects of zoom lens camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1105 –1110.

Li, S., Kwok, J. T., and Wang, Y. (2001). Combination of images with diverse focuses using the spatial frequency. *Information Fusion*, 2(3):169 – 176.

Liu, C., Szeliski, R., Kang, S. B., Zitnick, C., and Freeman, W. (2008). Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299 –314.

Liu, X., Yang, X., and Zhang, H. (2011). Fusion of depth maps based on confidence. In *International Conference on Electronics, Communications and Control*, pages 2658–2661.

Lobay, A. and Forsyth, D. (2006). Shape from texture without boundaries. *International Journal of Computer Vision*, 67:71–91.

Lorenzo, J., Castrillon, M., Mendez, J., and Deniz, O. (2008). Exploring the use if local binary patterns as focus measure. In *proc. International Conference on Computational Modeling, Control and Automation*, pages 855–860.

Mahmood, M. T. and Choi, T. S. (2012). Nonlinear approach for enhancement of image focus volume in shape from focus. *IEEE Transactions on image processing*, 21(5):2866–2873.

Mahmood, M. T., Choi, W. J., and Choi, T. S. (2008). PCA-based method for 3D shape recovery of microscopic objects from image focus using discrete cosine transform. *Microscopy Research and Technique*, 71(12):897–907.

Mahmoudi, M. and Sapiro, G. (2012). Sparse representations for range data restoration. *IEEE Transactions on Image Processing*, 21(5):2909–2915.

Malik, A. S. and Choi, T.-S. (2007). Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery. *Pattern Recognition*, 40(1):154–170.

Malik, A. S. and Choi, T.-S. (2008). A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise. *Pattern Recognition*, 41(7):2200–2225.

Malik, V., Cho, D., Shin, J., Har, D., and Paik, J. (2007). Color shift model-based segmentation and fusion for digital autofocusing. *Journal of Imaging Science and Technology*, 51(4):368–379.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674 –693.

Manjunath, B. and Ma, W. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837 –842.

Marshall, G. (2004). *Handbook of optical and laser scanning*. Marcel Dekker, Inc.

Marshall, J. A., Ariely, D., Burbeck, C. A., Aricly, T. D., Rolland, J. P., and Martin, K. E. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America*, 13:681–688.

Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London. B*, 263:169–172.

Mather, G. and Smith, D. R. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception*, 31:1211–1219.

Maybank, S. J. and Faugeras, O. D. (1992). A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8:123–151.

McLachlan, D. (1964). Extreme focal depth in microscopy. *Applied Optics*, 3(9):1009–1013.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. John Willey & Sons.

Melendez, J. (2010). *Supervised and unsupervised segmentation of textured images by efficient multi-level pattern classification*. PhD thesis, Universitat Rovira i Virgili.

Menn, N. (2004). *Practical Optics*. Elsevier Science Inc.

Minhas, R., Mohammed, A., and Wu, Q. (2012). An efficient algorithm for focus measure computation in constant time. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(1):152 –156.

Minhas, R., Mohammed, A. A., and Wu, Q. J. (2011). Shape from focus using fast discrete curvelet transform. *Pattern Recognition*, 44(4):839 – 853.

Minhas, R., Mohammed, A. A., Wu, Q. M., and Sid-Ahmed, M. A. (2009). 3D shape from focus and depth map computation using steerable filters. In *proc. International Conference on Image Analysis and Recognition*, pages 573–583, Berlin, Heidelberg. Springer-Verlag.

Montgomery, D. C. and Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Willey & Sons, 5th edition.

Muhammad, M. and Choi, T. S. (2011). An unorthodox approach towards shape from focus. In *IEEE International Conference on Image Processing*, pages 2965 –2968.

Muhammad, M. and Choi, T.-S. (2012). Sampling for shape from focus in optical microscopy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):564 –573.

Muhammad, M., Mutahira, H., Majid, A., and Choi, T. S. (2009). Recovering 3D shape of weak textured surfaces. In *International Conference on Computational Science and Its Applications*, pages 191 –197.

Muhammad, M. S. and Choi, T. S. (2010). A novel method for shape from focus in microscopy using bezier surface approximation. *Microscopy Research and Technique*, 73(2):140–151.

Murphy, D. B. (2001). *Fundamentals of light microscopy and electronic imaging*. Wiley-Liss.

Nabney, I. (2001). *NETLAB: Algorithms for Pattern Recognition*. Springer.

Nair, H. and Stewart, C. (1992). Robust focus ranging. In *proc. IEEE Conference on Computer Vision Pattern Recognition*, pages 309–314.

Namboodiri, V. P. and Chaudhuri, S. (2007). On defocus, diffusion and depth estimation. *Pattern Recognition Letters*, 28(3):311 – 319.

Nanda, H. and Cutler, R. (2001). Practical calibrations for a real-time digital omnidirectional camera. Technical report, Technical Sketches, Computer Vision and Pattern Recognition.

Navarro, R. (2009). The optical design of the human eye: a critical review. *Journal of Optometry*, 2:3–18.

Nayar, S. K. (1989). Shape from focus. Technical Report CMU-RI-TR-89-27, Carnegie Mellon University, Pitsburg, PA.

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

Nayar, S. K., Watanabe, M., and Noguchi, M. (1996). Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:1186–1198.

Ng, R. (2006). *Digital light field photography*. PhD thesis, Stanford Univesity.

Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. (2005). Light field photography with a hand-held plenoptic camera. Technical report, Standford University.

Nock, R. and Nielsen, F. (2004). Statistical region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1452 –1458.

Ojansivu, V. and Heikkila, J. (2007). Image registration using blur-invariant phase correlation. *IEEE Signal processing letters*, 14:449–452.

Ooi, K., Izumi, K., Nozaki, M., and Takeda, I. (1990). An advanced autofocus system for video camera using quasi condition reasoning. *IEEE Transactions on Consumer Electronics*, 36(3):526 –530.

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (1999). *Discrete-time digital signal processing*. Prentice Hall, 2nd edition.

Orieux, F., Giovanelli, J. F., and Rodet, T. (2010). Deconvolution with gaussian blur parameter and hyperparameters estimation. In *proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1350–1353.

Pahk, H. J., Lee, S. W., and Lee, D. S. (2000). Computer aided measurement and compensation system for focal length of lenses in camera manufacture based on the MTF performance using the line CCD sensor. *International Journal of Machine Tools and Manufacture*, 40(10):1493 – 1511.

Palou, G. and Salembier, P. (2013). Monocular depth ordering using t-junctions and convexity occlusion cues. *IEEE Transactions on Image Processing*, 22(5):1926–1939.

Paramanand, C. and Rajagopalan, A. N. (2012). Depth from motion and optical blur with an unscented Kalman filter. *IEEE Transactions on Image Processing*, 21:2798 – 2811.

Pech-Pacheco, J. L., Cristobal, G., Chamorro-Martinez, J., and Fernandez-Valdivia, J. (2000). Diatom autofocusing in brightfield microscopy: a comparative study. *proc. International Conference on Pattern Recognition*, 3:314–317.

Pelleg, D. and Moore, A. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *proc. International Conference on Machine Learning*, pages 727–734. Morgan Kaufmann.

Pentland, A. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531.

Pertuz, S., Garcia, M. A., and Puig, D. (2013a). Focus calibration for efficient focus sampling in conventional cameras. *IEEE Transactions on Image Processing*. (submitted).

Pertuz, S., Garcia, M. A., and Puig, D. (2013b). Modeling the focus profile in conventional cameras. *IEEE Transanctions on Pattern Analysis and Machine Intelligence*. (submitted).

Pertuz, S., Garcia, M. A., and Puig, D. (2013c). Shape estimation from autofocus. *Image and Vision Computing*. (submitted).

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

Pertuz, S., Puig, D., and Garcia, M. A. (2010). Improving shape from focus by compensating for image magnification shift. In *proc. International Conference on Pattern Recognition*, pages 802–805.

Pertuz, S., Puig, D., and Garcia, M. A. (2013e). Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432.

Petrou, M. and Sevilla, P. G. (2006). *Image Processing, Dealing with Texture*. John Willey & Sons.

Pieper, R. J. and Korpel, A. (1983). Image processing for extended depth of field. *Applied Optics*, 22(10):1449–1453.

Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. (2003). Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338 – 1351.

Pradeep, K. and Rajagopalan, A. (2007). Improving shape from focus using defocus cue. *IEEE Transactions on Image Processing*, 16(7):1920 –1925.

Pratt, W. K. (2007). *Digital Image processing: PISK scientific inside*. John Willey & Sons, 4th edition.

qing Qin, F. (2010). Blind image surper-resolution reconstruction based on PSF estimation. In *proc. International Conference on Information and Automation*, pages 1200–1203.

Raj, A. and Staunton, R. (2007). Estimation of image magnification using phase correlation. In *Int. Conference on Computational Intelligence and Multimedia Applications*, volume 3, pages 490–494.

Rajagopalan, A. and Chaudhuri, S. (1997). Optimal selection of camera parameters for recovery of depth from defocused images. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 219 –224.

Rajagopalan, A. and Chaudhuri, S. (1999). An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):577 –589.

Rao, A. R. and Lohse, G. L. (1996). Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, 36(11):1649 – 1669.

Reininger, R. and Gibson, J. (1983). Distributions of the two-dimensional DCT coefficients for images. *IEEE Transactions on Communications*, 31(6):835 – 839.

Remondino, F. and Fraser, C. (2006). Digital camera calibration methods: considerations and comparisons. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XXXVI, Dresden, Germany.

Rosete-Aguilar, M. (2000). Lens correction algorithms based on the see-saw diagram to correct seidel aberrations employing aspheric surfaces. In *International Conference on Education and Training on Optics and Photonics*, volume 3831, pages 412–419. SPIE.

Russell, M. J. and Douglas, T. S. (2007). Evaluation of autofocus algorithms for tuberculosis microscopy. In *proc. Anual International Conference of the IEEE Engeneering in Medicine and Biology Society*, pages 3489–3492.

Sahay, R. R. and Rajagopalan, A. N. (2008). A model-based approach to shape from focus. In *proc. International Conference on Computer Vision Theory and Applications*.

Samei, E., Flynn, M. J., and Reimann, D. A. (1998). A method for measuring the presampled MTF of digital radiographic systems using an edge test device. *Medical Physics*, 25(1):102–113.

Santos, A., de Solorzano, C. O., Vaquero, J. J., Pea, J. M., Mapica, N., and Pozo, F. D. (1997). Evaluation of autofocus functions in moleclar cytogenetic analysis. *Journal of Microscroscopy*, 188(3):264–272.

Sarkis, M., Senft, C. T., and Diepold, K. (2009). Calibrating and automatic zoom camera with moving least squares. *IEEE Transactions on Automation Science and Engineering*, 6(3):492–503.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 519–528.

Serrano, R. M. (2010). *Robust perceptual origanization techniques for analysis of color images*. PhD thesis, Polythecnic University of Catalunya.

Shen, C. and Chen, H. (2006). Robust focus measure for low-contrast images. In *Digest of Technical Papers of International Conference on Consumer Electronics*, pages 69–70.

Shim, S.-O. and Choi, T.-S. (2010). A novel iterative shape from focus algorithm based on combinatorial optimization. *Pattern Recognition*, 43(10):3338–3347.

Shim, S. O., Malik, A. S., and Choi, T. S. (2009). Accurate shape from focus based on focus adjustment in optical microscopy. *Microscopy Research and Technique*, 72:362–370.

Shirai, K. and Ikehara, M. (2005). All-in-focus photo image creation by wavelet transform. In *Asilomar Conference on Signals, Systems and Computers*, pages 888 – 892.

Shirvaikar, M. (2004). An optimal measure for camera focus and exposure. In *proc. Southeastern Symposium on System Theory*, pages 472 – 475.

Shoji, H., Shirai, K., and Ikehara, M. (2006). Shape from focus using color segmentation and bilateral filter. In *4th - Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop*, pages 566–571.

Sikora, T. (1997). MPEG digital video-coding standards. *IEEE Signal Processing Magazine*, 14(5):82 –100.

Stokseth, P. A. (1969). Properties of a defocused optica-system. *Journal of the Optical Society of America*, 59:1314–1321.

Strang, G. and Nguyen, T. (1996). *Wavelets and filter banks*. Wellesley-Cambridge Press, 2nd edition.

Sturm, P. and Maybank, S. (1999). On plane-based camera calibration: A general algorithm, singularities, applications. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 432–437.

Subbarao, M. (1988). Parallel depth recovery by changing camera parameters. In *proc. International Conference on Computer Vision*, pages 149 –155.

Subbarao, M. and Choi, T. (1995). Accurate recovery of three-dimensional shape from image focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):266–274.

Subbarao, M., Choi, T., and Nikzad, A. (1993). Focusing techniques. *Journal of Optical Engineering*, 32:2824–2836.

Subbarao, M. and Gurumoorthy, N. (1988). Depth recovery from blurred edges. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 498 –503.

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

Subbarao, M. and Tian, J. K. (1998). Selecting the optimal focus measure for autofocusing and depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):864–870.

Sugimoto, S. A. and Ichioka, Y. (1985). Digital composition of images with increased depth of focus considering depth information. *Applied Optics*, 24(14):2076–2080.

Sun, Y., Duthaler, S., and Nelson, B. J. (2004). Autofocusing in computer microscopy: Selecting the optimal focus algorithm. *Microscrocopy Research and Technique*, 65(3):139–149.

Sundaram, H. and Nayar, S. (1997). Are textureless scenes recoverable? In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 814–820.

Szeliski, R. (2011). *Computer vision: algorithms and applications*. Springer.

Taubman, D. S. and Marcellin, M. W. (2002). *JPEG 2000: image compression fundamentals, standards and practice*. Kluwer Academic Publishers.

Tay, C., Thakur, M., Chen, L., and Shakher, C. (2005). Measurement of focal length of lens using phase shifting Lau phase interferometry. *Optics Communications*, 248(4-6):339 – 345.

Taylor, J. R. (1997). *An introduction to error analysis. The study of uncertainties in physical measurements*. University Science Books, 2nd edition.

Tenenbaum, J. M. (1971). *Accommodation in computer vision*. PhD thesis, Stanford University.

Thelen, A., Frey, S., Hirsch, S., and Hering, P. (2009). Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. *IEEE Transactions on Image Processing*, 18(1):151–157.

Thibos, L. N. (2009). Retinal image quality for virtual eyes generated by a statistical model of ocular wavefront aberrations. *Ophtalmic and physiological optics*, 29:288–291.

Tian, J. and Chen, L. (2010). Multi-focus image fusion using wavelet-domain statistics. In *proc. IEEE International Conference on Image Processing*, pages 1205 –1208.

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

Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *proc. International Conference on Computer Vision*, pages 839 –846.

Torre, V. and Poggio, T. A. (1986). On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):147 –163.

Tovee, M. J. (2008). *An introduction to the visual system*. Cambridge University Press.

Triesch, J., Ballard, D. H., and Jacobs, R. A. (2002). Fast temporal dynamics of visual cue integration. *Perception*, 31:421–434.

Tsai, D. C. and Chen, H. H. (2012). Reciprocal focus profile. *IEEE Transactions on Image Processing*, 21(2):459 –468.

Tsai, R. Y. (1986). An efficient and accurate camera calibration technique for 3D machine vision. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374.

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323 –344.

Ullman, S. (1979). The interpretation of the structure from motion. *Proceedings of the Royal Society of London*, 203:405–426.

Vaquero, D., Gelfand, N., Tico, M., Pulli, K., and Turk, M. (2011). Generalized autofocus. In *IEEE Workshop on Applications of Computer Vision*, pages 511 –518.

Vaughnn, D. and Mark, R. (2006). Actual field curvature and isoquals. In Mouroulis, P. Z., Smith, W. J., and Johnson, R. B., editors, *SPIE Proceedings, Current Developments in Lens Design and Optical Engineering VII*, volume 6288, page 62880E.

Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., and Tumblin, J. (2007). Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*, 26(3).

Voelz, D. G. (2010). *Computational Fourier Optics*. SPIE Press.

Vogel, C. R. and Oman, M. (1998). Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Transactions on Image Processing*, 7(6):813–824.

von Helmholtz, H. (1924). *Helmholtz's Treatise of Physiological Optics*. The Optical Society of America.

Wallace, G. (1992). The JPEG still picture compression standard. *IEEE Transactions on Consumer Electronics*, 38(1):xviii –xxxiv.

Wallach, H. and O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45:205–217.

Wang, W., Wang, H., Hempel, M., Peng, D., Sharif, H., and Chen, H.-H. (2011). Secure stochastic ECG signals based on Gaussian mixture model for e-healthcare systems. *IEEE Systems Journal*, 5(4):564 –573.

Wang, W. W., Shui, P. L., and Song, G. X. (2003). Multifocus image fusion in wavelet domain. In *proc. International Conference on Machine Learning and Cybernetics*, volume 5, pages 2887–2890.

Wang, Z. and Bovik, A. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81 –84.

Watanabe, M. and Nayar, S. (1995). Telecentric optics for constant magnification imaging. Technical report, Dept. of Computer Science, Columbia University.

Watanabe, M. and Nayar, S. K. (1998). Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27:203–225.

Watson, A. B. and Ahumada, A. J. (2008). Predicting the visual acuity from wavefront aberrations. *Journal of Vision*, 8(4):1–19.

Wee, C. and Paramesran, R. (2007). Measure of image sharpness using eigenvalues. *Information Sciences*, 177(12):2533 – 2552.

Wee, C. Y. and Paramesran, R. (2008a). Comparative analysis of eigenvalues-based and Tchebichef moments-based focus measures. In *International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 1, pages 517–520.

Wee, C. Y. and Paramesran, R. (2008b). Image sharpness measure using eigenvalues. In *International Conference on Signal Processing*, pages 840–843.

Weng, J., Cohen, P., and Hermiou, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980.

Wheatstone, C. (1838). Contributions to the physiology of vision. *Philosophy transactions of the Royal Society of London*, 128:371–394.

Williams, T. L. (1999). *The optical transfer function of imaging systems*. CRC Press.

Willson, R. G. (1994). *Modeling and calibration of automated zoom lenses*. PhD thesis, Carnegie Mellon University.

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, 17:17–45.

Wolberg, G. (1990). *Digital Image Warping*. IEEE Computer Society Press.

Wrigley, S., Brown, G., Wan, V., and Renals, S. (2005). Speech and crosstalk detection in multichannel audio. *IEEE Transactions on Speech and Audio Processing*, 13(1):84 – 91.

Wu, Q. (2008). Chapter 16 - autofocusing. In *Microscope Image Processing*, pages 441 – 467. Academic Press, Burlington.

Xian, T. (2006). *Three-dimensional modeling and autofocusing technology for new generation digital cameras*. PhD thesis, Stony Brook University.

Xie, H., Rong, W., and Sun, L. (2006). Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images. In *proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 229–234.

Xie, H., Rong, W., and Sun, L. (2007). Construction and evaluation of a wavelet-based focus measure for microscopy imaging. *Microscopy Research and Technique*, 70(11):987–995.

Xiong, Y. and Shafer, S. (1993). Depth from focusing and defocusing. In *proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 68 –73.

Yang, G. and Nelson, B. (2003). Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *proc. International Conference on Intelligent Robots and Systems*, volume 3, pages 2143–2148.

Yao, Y., Abidi, B., and Abidi, M. (2006). Digital imaging with extreme zoom: System design and image restoration. In *proc. IEEE International Conference on Computer Vision Systems*, pages 52–58.

Yap, P. and Raveendran, P. (2004). Image focus measure based on Chebyshev moments. In *IEE proc. Vision, Image and Signal Processing*, volume 151, pages 128–136.

Yates, R. and Goodman, D. (1999). *Probability and Stochastic processes*. John Willey & Sons.

Yousefi, S., Rahman, M., and Kehtarnavaz, N. (2011). A new auto-focus sharpness function for digital and smart-phone cameras. *IEEE Transactions on Consumer Electronics*, 57(3):1003 –1009.

Yun, J. and Choi, T. (1999). Accurate 3-d shape recovery using curved window focus measure. In *proc. International Conference on Image Processing*, volume 3, pages 910–914 vol.3.

Zerene Systems (2011). Zerene stacker. retrieved on 01/04/2011 from http://www.zerenesystems.com/cms/stacker.

Zhang, R., Tsai, P. S., Cryer, J. E., and Shah, M. (1999). Shape from shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:690–706.

Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. In *proc. IEEE International Conference on Computer Vision*, volume 1, pages 666 –673.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 22(11):1330 – 1334.