Ecole Doctorale Environnement - Santé / STIC (E2S)

A dissertation

by

Intuon Lertrusdachakul

presented to the Université de Bourgogne in partial fulfillment of the requirements for the degree of

Docteur de l'Université

discipline Instrumentation et Informatique de l'Image

A Novel 3D recovery method by dynamic (de)focused projection

---

30 November 2011

Jury

François GOUDAIL	Professor	Reviewer
Majdi KHOUDEIR	Professor	Reviewer
Ernest IRSCH	Professor	Examiner
Anastasia ZAKHAROVA	Associate Professor	Examiner
Yohan FOUGEROLLE	Associate Professor	Advisor
Olivier LALIGANT	Professor	Ph.D Director

Acknowledgement

First and most important, I would like to express my sincere and deep gratitude to my Ph.D director, Professor Olivier Laligant, for his valuable guidance and constant academic support during my Ph.D. research. I enjoyed working and learned a tremendous amount under his supervision. I also thank my advisor, Associate Professor Yohan Fougerolle. The discussions with him have been very motivational.

I truly appreciate my committee members, Professor François Goudail, Professor Majdi Khoudeir, Professor Ernest Irsch, and Associated Professor Anastasia Zakharova, for their precious time and valuable suggestions on this dissertation.

My thanks also goes to all the Le2i members and our lovely secretary (Nathalie) for their precious friendships and helpfulness during this wonderful 3 years. I give special thanks in particular to Pierre-Emmanuel Leni, Youssef Bokhabrine, and Romuald Jolivot, for their inspiration and technical support.

Last but not least, I would express my deepest gratitude and thanks to Lertrusdachakul family for their continuous support and patiently understanding. Their endless encouragement and support made this dissertation possible. I cannot thank them enough for their unconditional love.

The work in this thesis were supported by Ministère de l'enseignement supérieur et de la recherche (MESR). It is also gratefully acknowledged.

Résumé

Ce mémoire présente une nouvelle méthode pour l'acquisition 3D basée sur la lumière structurée. Cette méthode unifie les techniques de *depth from focus* (DFF) et *depth from defocus* (DFD) en utilisant une projection dynamique (dé)focalisée. Avec cette approche, le système d'acquisition d'images est construit de manière à conserver la totalité de l'objet nette sur toutes les images. Ainsi, seuls les motifs projetés sont soumis aux déformations de dé-focalisation en fonction de la profondeur de l'objet. Quand les motifs projetés ne sont pas focalisés, leurs *Point Spread Function* (PSF) sont assimilées à une distribution gaussienne. La profondeur finale est calculée en utilisant la relation entre les PSF de différents niveaux de flous et les variations de la profondeur de l'objet. Notre nouvelle estimation de la profondeur peut être utilisée indépendamment. Elle ne souffre pas de problèmes d'occultation ou de mise en correspondance. De plus, elle gère les surfaces sans texture et semi-réfléchissante. Les résultats expérimentaux sur des objets réels démontrent l'efficacité de notre approche, qui offre une estimation de la profondeur fiable et un temps de calcul réduit. La méthode utilise moins d'images que les approches DFF et contrairement aux approches DFD, elle assure que le PSF est localement unique.

Abstract

A Novel 3D recovery method by dynamic (de)focused projection

This paper presents a novel 3D recovery method based on structured light. This method unifies depth from focus (DFF) and depth from defocus (DFD) techniques with the use of a dynamic (de)focused projection. With this approach, the image acquisition system is specifically constructed to keep a whole object sharp in all of the captured images. Therefore, only the projected patterns experience different defocused deformations according to the object's depths. When the projected patterns are out of focus, their Point Spread Function (PSF) is assumed to follow a Gaussian distribution. The final depth is computed by the analysis of the relationship between the sets of PSFs obtained from different blurs and the variation of the object's depths. Our new depth estimation can be employed as a stand-alone strategy. It has no problem with occlusion and correspondence issues. Moreover, it handles textureless and partially reflective surfaces. The experimental results on real objects demonstrate the effective performance of our approach, providing reliable depth estimation and competitive time consumption. It uses fewer input images than DFF, and unlike DFD, it ensures that the PSF is locally unique.

Keywords: focus, depth from defocus, active illumination pattern, range sensor, blur estimation, 3D reconstruction

Contents

Chapter 1

Introduction

In the extensive field of computer vision, depth recovery approaches have been broadly developed and have attracted substantial attention over recent decades. It is a challenging problem to recover the 3D information (i.e., depth) that is lost during the projection of a 3D scene onto a 2D image plane. Several 3D reconstruction algorithms have already been proposed; the effectiveness of each method, both qualitative and quantitative, has been studied. However, there is still no unique satisfactory solution that applies to all kind of scenes. Moreover, our new approach, which entails a combination of depth from focus (DFF) and depth from defocus (DFD) with the use of a light pattern projection, has not yet been explored. In this thesis, we develop a prototype of the range sensor from our new depth estimation system.

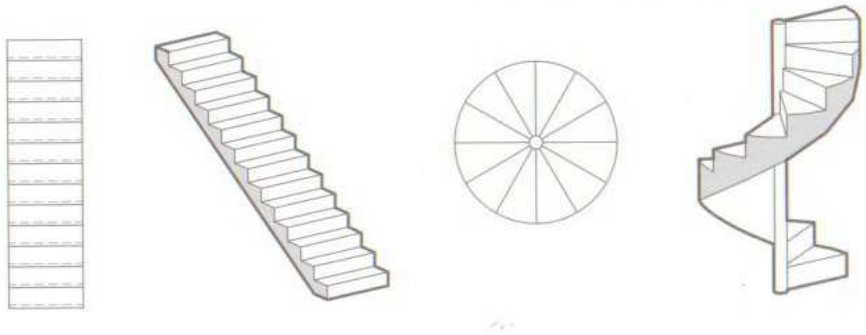
1.1 Role of 3D reconstruction in computer vision

Nowadays, many advanced applications require 3D data. The depth information plays an important role in the analysis of dynamic or static environments. It is the fundamental clue for other relevant machine vision and artificial intelligence applications such as recognition, classification, modeling. Real world applications also include the surveillance and robotic field that exploit depth information to gain better environmental analysis. In the medical

field, new technologies and researches also applied more and more reliable depth data. Broadly speaking, the domains of 3D image processing, digital photography, multimedia, 3D visualization and augmented reality have increasingly utilized 3D information. For our specific system, the prototype is developed for a small object and has ability to deal with difficult surface textures. In addition, the system uses video projector as a light source. Therefore, live objects (animal and plant) can be safely tested. Other particular examples can be found in biological specimen analysis, and defect metallic component detection for industrial purpose.

1.2 Motivations and Objectives

Different scenes may appear identical from certain viewpoints. Figure 1.1 shows straight run stair and spiral stair with similar steps appearance, returning same top view images. Clearly, it is difficult to distinguish such scenes without depth information.

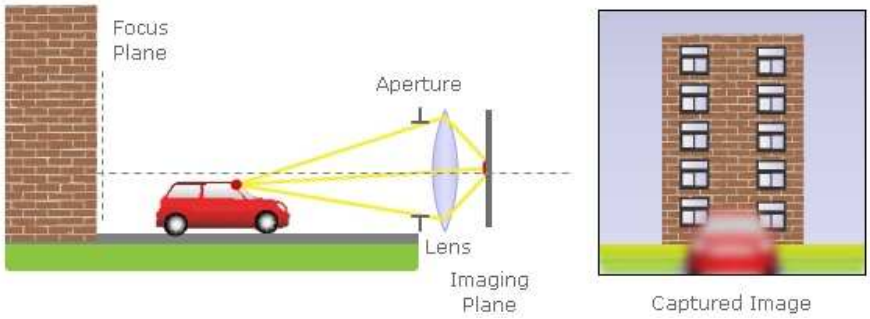


(a) (b)

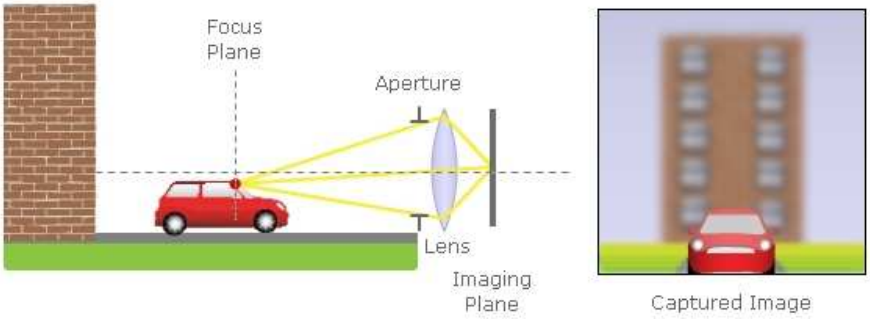
Figure 1.1: Top views and 3D views of (a) Straight run stair (b) spiral stair

In general, there are 2 types of 3D data to be considered: photometric and geometric information. The 3D shapes and distances of objects in the scene constitute the geometric information (i.e., depth map), which is also related to the spatial distribution occupied by the object. While, light energy of the image constitutes the photometric information (i.e., brightness).

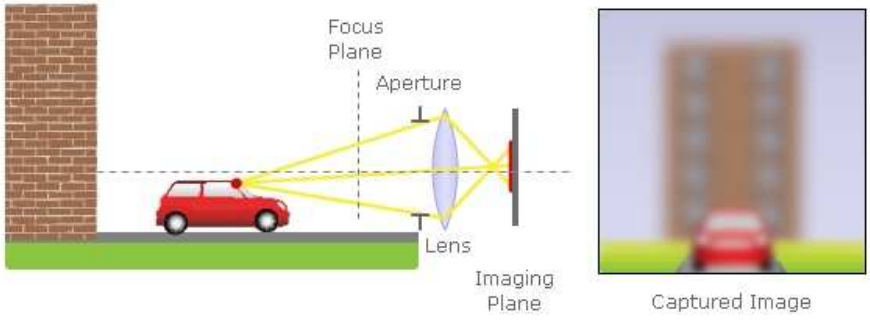
For a given optical setting, scene points lying on focal plane located at certain distance from the lens will be correctly focused on the sensor. Only finite range of the distances called depth of field (DOF) are acceptably sharp. While, scene points at further distances away from the focal plane will appear increasingly blurred due to the defocus. This suggests that the defocus and the geometry are related. Therefore, it is possible to estimate the geometry of a scene by measuring the blur levels in an image. However, we cannot make the distinction between the blurred image of an object with sharp texture, and the sharp image of an object with blurred texture. Alternatively, when two scenes are defocused at different degrees, they can produce identical images. Hence, a single defocused image is not sufficient to obtain a unique reconstruction of the scene unless additional information is available. To deal with this problem, we need to analyze two or more defocused images obtained with different focus settings, as shown in figure 1.2. Four images of the same scene taken with different camera settings. (a) The building is brought into focus. (b) The car is brought into focus. (c) and (d) Nothing is brought into focus. Both the building and the car are defocused. In addition, the building is more blurred than the car since it is farther from the focus plane. Therefore, when the images are taken at various amount of blurs, we can quantify the blur levels (e.g., size of blur, defocus measure) which can be eventually used for depth estimation and 3D reconstruction, respectively.



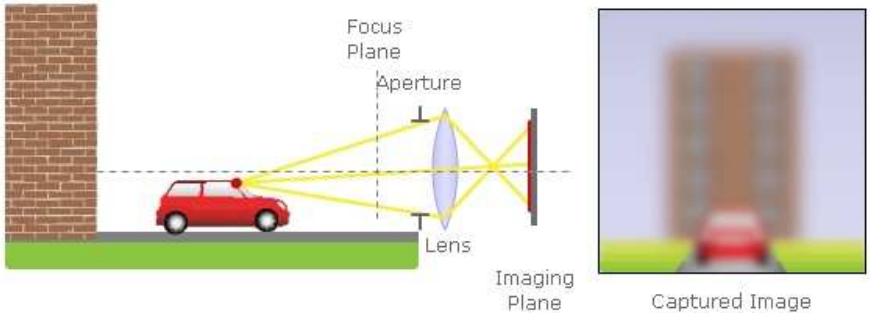
(a)



(b)



(c)



(d)

Figure 1.2: Scene with different camera settings [1]

Our aim is to introduce a new and alternative 3D reconstruction system that can recover the depth of objects from a collection of defocused images, and solve some specific problems found in classical approaches such as weak texture surface, occlusion problem, and correspondence problem. We focus on the novelty, uniqueness, and feasibility of the system.

1.3 Challenges in depth retrieval

There are numbers of conditions either the object itself or the system environment that create difficulties to the depth recovery system (see figure 1.3). Some of the prominent challenges in this area are listed below:

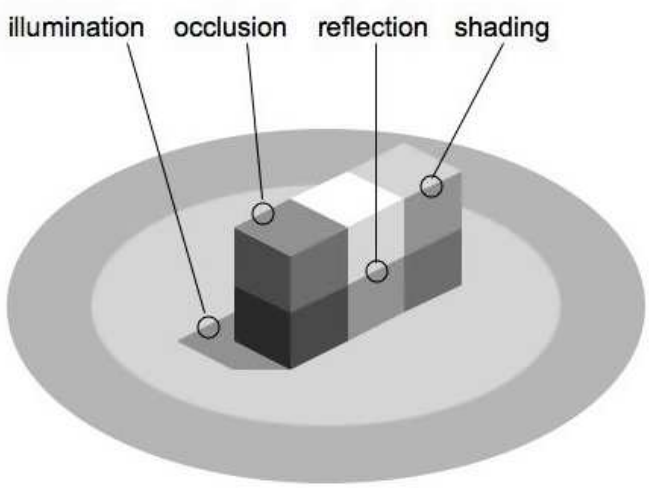


Figure 1.3: Some prominent challenges in depth retrieval

Occlusion. It is one of the major problems that occurs when surface or parts of the surface is not visible from a certain viewpoint. In addition, as the aspect changes, one part may be occluded by another part of the same object (self-occlusion).

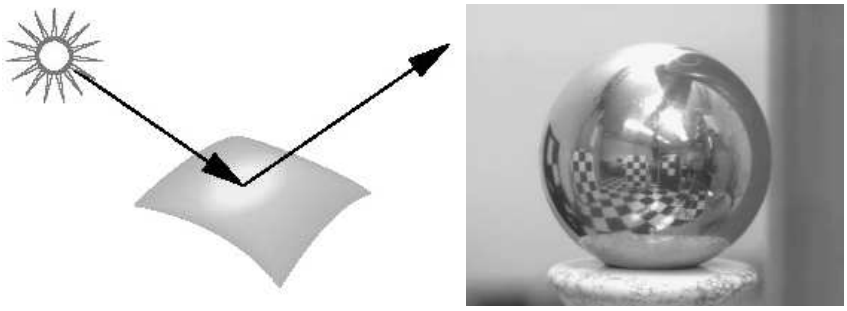
Correspondence matching. Correspondence matching and feature tracking are the problem found in traditional approaches such as stereo and shape from motion, and are time consuming.

Lack of texture. A weak textured surface does not provide sufficient details for depth estimation because both focus and defocus give the same representation. Example of textureless objects is illustrated in figure 1.4.

Figure 1.4: Example of textureless objects

Transparent and dark surfaces. When an object surface totally or partially pass/absorb the light, it is difficult to identify the object or to infer the shape from acquired images.

Reflectance. High reflectivity can lead to specularity problem (see figure 1.5), in both normal image based algorithms and structured light based algorithms.



(a) (b)

Figure 1.5: Reflectance problem (a) Specularity problem, and (b) highly specular surface

Illumination, scale variation. Variation in environmental illumination causes large variations in the intensity values of pixels. Some related common issues are formation of shadow, non-linear transformation in pixels, scaling, and shifting due to the change in position of light source.

Noise. Noise can degrade the image by causing some information loss.

We aim to develop practically usable system, that can address most of these challenges. Moreover, the successful approach to the object reconstruction should also emphasize on minimizing financial system cost. While, making it robust, accurate, and computationally efficient.

1.4 Contribution of the thesis

In this thesis, we have proposed a new solution to the 3D recovery system using dynamic defocused projection. Short note on each of the contributions is given below:

The demonstration of a novel 3D retrieval approach. The first contribution is the demonstration of a novel 3D retrieval approach based on active structured light. We address our methods as the combination of two related 3D recovery approaches, which are depth from focus (DFF) and depth from defocus (DFD). We formulate a principle and prototype of the new depth estimation by analyzing the point spread function (PSF) of the defocus model. This study provides the results in both qualitative and qualitative aspects. However, the decisive goal of this work is to verify the novelty, uniqueness, and feasibility of the system.

The investigation of the benefits of this approach over other 3D recovery methods. Our approach can be employed as a stand-alone strategy returning effective dense depth maps. The method has tested on the real objects that has textureless surface and has some reflectivity. By projecting the illumination pattern, the method can overcome the problem of weak texture, which is the common bottleneck among all passive 3D reconstruction algorithms. Moreover, it does not suffer neither from the correspondence problem nor the occlusion problem found in traditional approaches. Our system is developed specifically for small object reconstruction. However, the scale of the setup can be enlarged without altering the system principle.

The derivation of new depth estimation. The system prototype is derived from our new depth estimation. We construct the specific system, in which the PSF is locally unique. It is composed of 3 main subsystems: the light pattern projection system, the main optical system, and the acquisition system. The projecting light patterns are also carefully designed for the distortion optimization. The system requires the calibration to be done only once for each type of object material. The image acquisition, the image profile analysis, the pattern localization, and the spread parameter calculation are performed for both calibrated objects and test objects to obtain the blur estimates. Eventually, the depth is evaluated from the analysis of the relationship between the sets of PSFs obtained from different blurs and the variation of the object's depths.

1.5 Organization of the thesis

The remainder of the thesis is structured as follows. In chapter 2, we have summarized the detailed survey of the literatures found in shape recovery field especially for non-contact based methods. Next, we focus on the depth from focus and depth from defocus method. Some related works and the fundamental background of camera geometry and the blur properties are explained in Chapter 3. Chapter 4 provides the prototype principle, the system components, and the depth estimation. We describe the implementation and obtained experimental results in Chapter 5. In the last chapter, we conclude by summarizing the performance expectations and limitations of this new technique and also discussing further possible research perspectives.

Chapter 2

Literature Review

One of the goals of computer vision is to predict the shape of the 3D scene by extracting useful information from 2D images. Either absolute depth or relative depth can be recovered. The absolute depth cues (i.e., distance perception) can be used to determine the exact spatial location of an object or physical distance. It is based primarily on the monocular depth cues. While, the relative depth cues (i.e., depth perception) can indicate which objects are closer or further with respect to other objects. It is based on monocular and binocular depth cues.

A well established classification divides shape recovery approaches into 2 types: contact and non-contact methods (see figure 2.1). Contact-based methods generally provide high accuracy and are suitable for a wide range of the applications. However, these methods involve mechanical movements of a probe device from one measurement point to another. Consequently, the data acquisition can be very time consuming. There is also the disadvantage of having to be in contact with the objects. It is sometimes inadmissible particularly very soft object. Example of device using contact-based method is the Coordinate Measuring Machine (CMMs). We often find this digitization systems using the lasers mounted on some sort of the jointed arms with the high degree of freedom. These arms can either be operated manually or automatically. Through their internal positioning system, the geometry information is transferred to the controlling software. The machine are very large in size and capable for measuring large object dimensions. They require

the spacious platform and need to be placed in protected environment rooms against temperature variation and vibrations.

The current development of 3D modeling is therefore mainly achieved by using non-contact based methods.

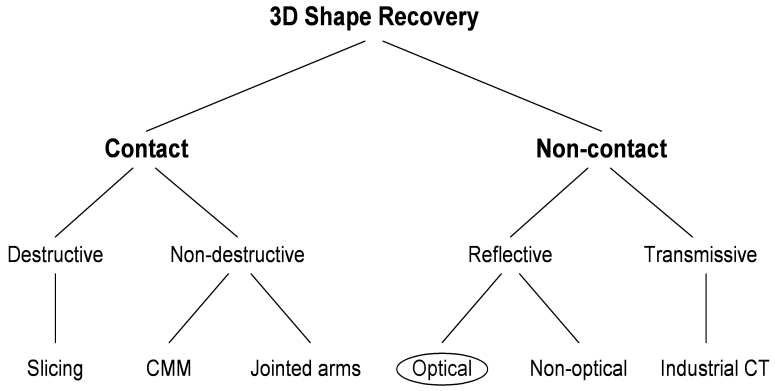


Figure 2.1: Hierarchy of shape acquisition

2.1 Non contact depth recovery methods

Here, we clarify 2 possible ways to classify non-contact 3D recovery approaches:

1. 2 Classes: Passive and active techniques or
2. 3 Classes: Time delay based, triangulation based, and imaging cue based.

There are a number of methods that fall under each of these categories. In this chapter, we first give a brief overview of the previous related works in the optical field. More details emphasized on the focus/defocus and our approach are provided in the latter chapter.

2.1.1 Passive and active techniques

As seen from figure 2.2, within the optical field, there is a clear distinction between passive and active techniques whether some sort of energy projection (e.g. lasers, structured

light) or the relative motion between camera(s) and objects are considered or not. Active techniques (e.g. laser, sonar ranging) use either laser beam or sound wave to sense the shape of the scene objects. While, passive techniques work under ambient lighting to acquire data for 3D recovery. Both methods have advantages and disadvantages.

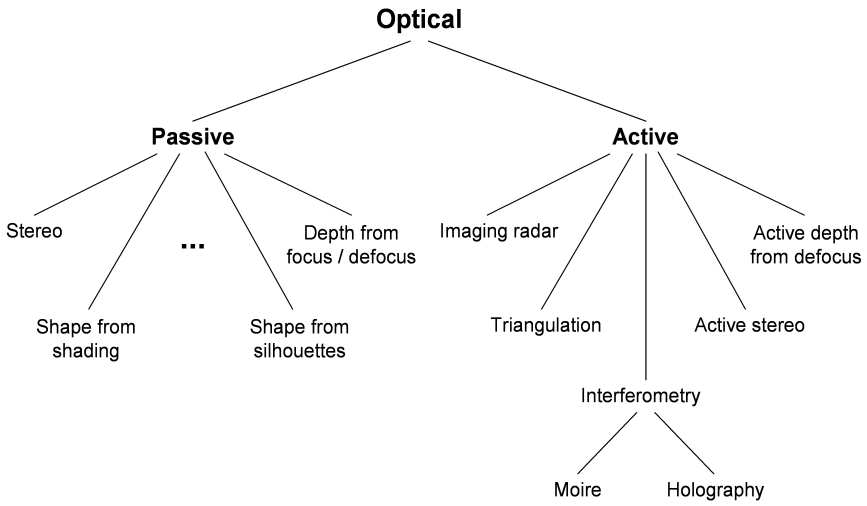


Figure 2.2: Hierarchy of 3D reconstruction methods in the optical field

Passive techniques, such as stereo and shape from motion, use at least two images to perform multiple view correspondence matching [2], [3]. The depth is extracted from either the disparity or motion vectors after matching. The main drawbacks of these techniques are that they are computationally expensive to either perform correspondence matching or feature tracking as well as the occlusion problems in scene areas that are visible only by one camera [4]. Other passive techniques include shape from shading and shape from texture. By using only a single image, the depth ambiguities can be retrieved. However, these techniques are only complementary to other strategies. Overall, the common bottleneck shared among all of the passive techniques is that the depth cannot be computed accurately in the case of weak texture or textureless scenes [5]. More details with references of these popular passive methods can be found in Section 2.1.2. Meanwhile, active techniques use active illumination to solve texture problems and are generally based on the principle of structured light. The most well-known active techniques are the Light striping method, Moiré interferometry, and Fourier-transform profilometry [6],[7]. Depth can be extracted from the image deformation of projected

pattern [8],[9]. Nevertheless, Moiré interferometry and Fourier-transform profilometry return only a relative depth, not an absolute depth.

Overall, the active technique usually requires specific hardware and setting to produce accurate and reliable estimation. In contrast, the passive technique has relatively lower system cost and less complexity. However, the accuracy and global performance may not be as good as the active methods in some aspects.

2.1.2 Time delay, Triangulation, Imaging cue based Methods

In this section, we propose alternative strategy to categorize the non-contact depth recovery methods. Figure 2.3 shows a hierarchical representation of the existing approaches.

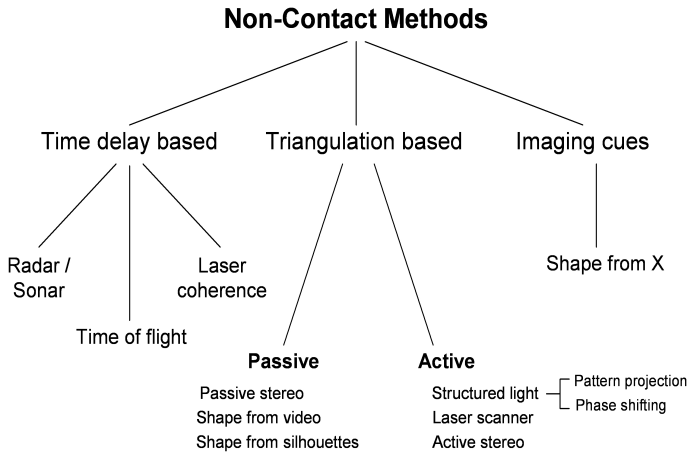


Figure 2.3: Hierarchy of non-contact 3D recovery approaches

Time delay based

The first category is called time delay based approaches, where a transreceiver system computes the delay or any deterioration in reflected signal after sending and returning back from object surfaces. The elapsed time between emission and detection yields the distance to the object since the speed of the laser light is precisely known.

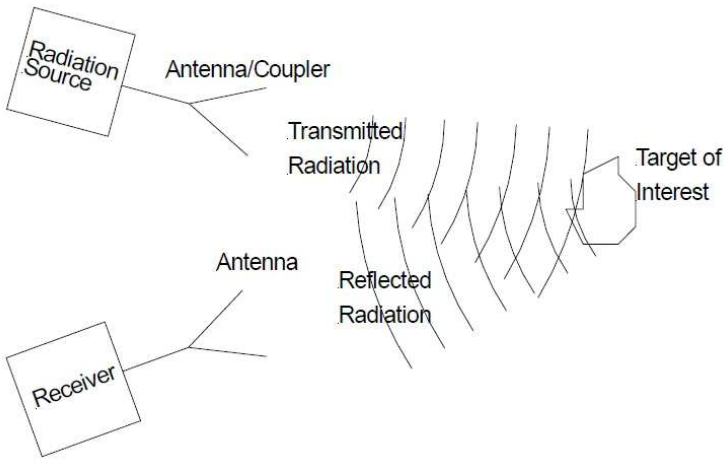


Figure 2.4: The operational principles of a generic time-of-flight sensor

Time-of-flight is the principle mode of operation for most radar, laser and active acoustic devices. The operational principles of a generic Time-of-flight sensor is illustrated in figure 2.4. This technique measures the time between the transmission of a pulse and the reception of an echo to provide range according to the following equation:

$$R = \frac{v\Delta T}{2} \tag{2.1}$$

where

$$\begin{cases} R & \text{is the range (m)} \\ v & \text{is the wave propagation velocity (m/s)} \\ \Delta T & \text{is the round trip time (s)} \end{cases}$$

Because the round-trip time is calculated, there is a factor of 2 in the formula as shown in equation 2.1. To operate efficiently, a narrow beam must be formed to concentrate the transmitted energy, the transducer must be matched to the characteristics of the medium, and the receiver must match the transmitter characteristics [10]. The sensors operational principles are the same for electromagnetic (e.g. radar, laser) and active acoustic sensing. Such method provides useful rough depth maps for extensive distance scene but the scanning time can be very long. Example of pulsed time-of-flight radar is shown in figure 2.5, where the echoes from the nearer houses return to the radar first, and the echoes from the last two houses overlap and appear to the radar as a single return.

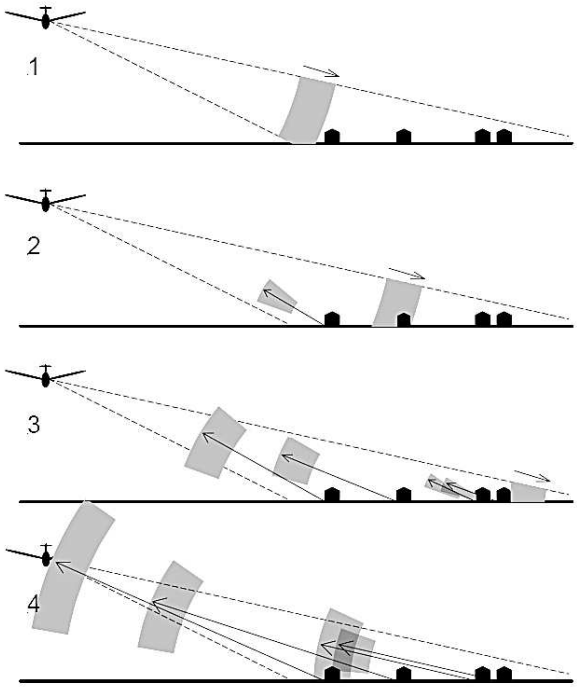


Figure 2.5: The operational principles of a pulsed time-of-flight radar

Triangulation based

The second class of approaches uses a geometric formulation to infer depth, known as triangulation. Shape from video, shape from silhouettes, structured light, and traditional laser scanner are example of this approach. Some prominent examples are discussed below:

Laser scanning. A system consists of a laser source and an optical detector. The laser source emits light in the form of either a line or a pattern on the object surface. While, the optical detector (i.e., digital camera) is used to detect this line or pattern on the objects. By applying the triangulation principle as shown in figure 2.6, the system is able to deduce the distance from the objects and extract their geometry, accordingly. The advantage of using laser sources is that laser light is very bright and highly focused for long distances. As a result, the emitted pattern is always focused on the surface of the objects. These laser scanning techniques are typically used independently, but can also use in combination to create a more versatile scanning system. Moreover, it provides high accuracy in geometry

measurements. The main drawback is the high cost of such system, which render this method useful to only certain applications. Additionally, special attention should be paid for surfaces with specific properties, such as reflectance and transparency. Lastly, in terms of productivity and portability, it can vary significantly [11], [12].

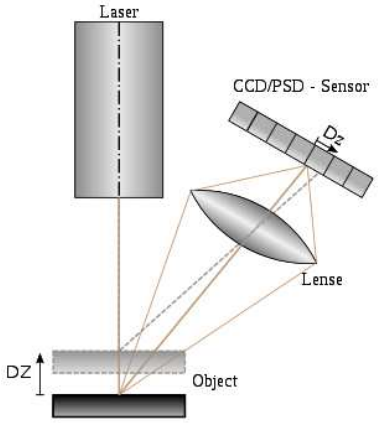


Figure 2.6: The operational principles of a laser scanner

Structured Light. This method is also based on triangulation, but does not need to use specific laser sources. As seen in figure 2.7, the method works by projecting a predefined light pattern to cover the whole/partial surface of the objects. The scene is then captured by a detector and processed in order to deduce the geometry from the pattern deformations in the digital image. These patterns can be simple multiple fringes of different colors or complex patterns with curves, either time or space coded [13],[14]. The method is accompanied by texture acquisition and can lead to high accuracy and productive results. Moreover, the systems are usually portable and easy to use.

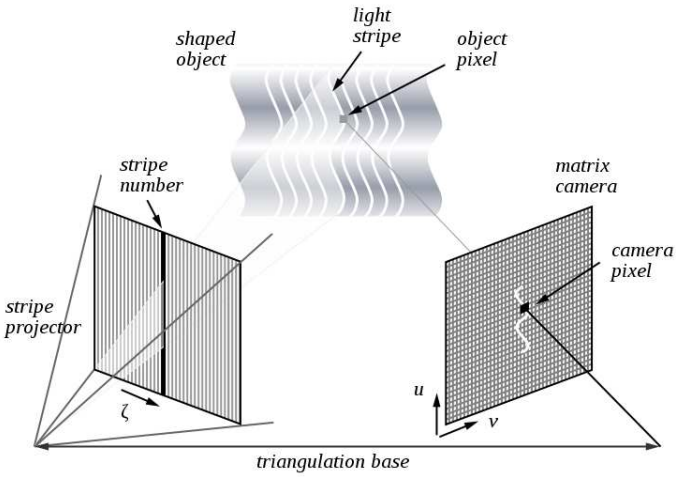
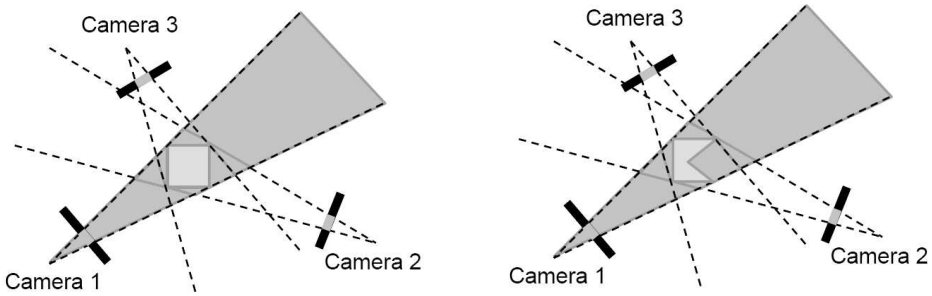


Figure 2.7: The operational principles of a Structure light scanner

Shape from silhouette. This technique recovers the object shape maximally consistent with the silhouettes (i.e., the visual hull) [15], [16]. It is basing on multiple photographic capturing of the object from different viewing angles. Then, we can deduce the geometry from their object's silhouettes. Due to its simplicity and the speed, it is often used in the real-time system. Moreover, the automated process can be high productivity and low cost. It can capture both geometry and texture. The main disadvantage is that it cannot recover concavities (that is not visible in the silhouettes). Figure 2.8 illustrates the operational principle and its concavity issue.



(a) (b)

Figure 2.8: Shape from silhouette method (a) and its concavity problem (b)

Shape from video. This method is a variant of shape from stereo. Two photo cameras in stereo vision are replaced by a video camera. It captures the object in a sequence of

images from different views. A requirement for the system is that the object need to be fixed completely with no movable parts. A key point of the algorithm is the identification of common points between different images and the registration of these points onto a virtual 3D scene. There is no prior knowledge about the position of the camera or the objects. Advantages of this method are the portability, the ability to capture both texture and geometry, and also relatively low cost. Significant disadvantage is the low resolution in capturing the geometry and sensitivity of noise in video sequences [17].

Imaging cues

The last category is based on imaging cues. It is also known as shape from X approaches, where X can be stereo, texture, shading, motion and defocus. Some detailed description of these approaches are summarized below:

Shape from Stereo. Stereo imaging depicts how the human eyes works. The system uses a pair or more number of stereo images to calculate the relative displacement (i.e., disparity) between the image features in the frames. The basic stereo system principle is illustrated in figure 2.9. Since the disparity is inversely proportional to the depth, this leads to depth estimation. The major difficulty of this technique is to obtain accurate correspondence between the image features. The image features can be either sparse (e.g. Harris corners, SIFT features) or dense based in pixel intensity values at each pixel location. We can classify correspondence algorithms into correlation-based and feature-based algorithms. In correlation-based algorithms, the elements to match are image windows of fixed size, and the similarity criterion is a measure of the correlation between the windows in the two images. These algorithms typically give dense measurements of depth. On the other hand, feature-based methods use a set of features to find correspondence in two images. The distance between feature descriptors is measured with the numerical and symbolical properties of the features. Corresponding elements are given by the most similar feature pair. Other related works solve the correspondence problem by formulating an error function which can be solved using dynamic programming [18], [19], graph based

methods [20], belief propagation [21], and simulated annealing [22]. In multi-view stereo, there are multiple views of the scene to obtain the three-dimensional information [23], [24], [25]. The output of shape from stereo is the disparity map or depth map.

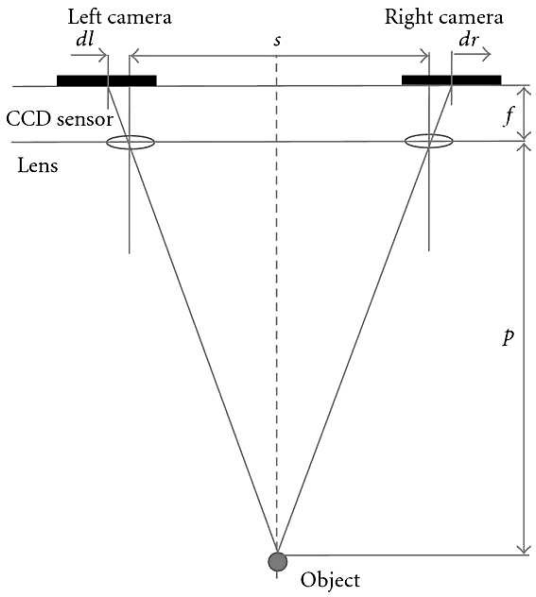


Figure 2.9: The basic stereo vision

Shape from Shading (SFS). Figure 2.10 shows the principle of shape from shading. It uses a gradual variation of shading in an image to infer the shape of the surface. The key idea is the relation between the image intensity and the surface slope. The radiance on each image point is used to calculate the surface normal, the direction of the illumination pointing towards the light source, and the albedo of the surface. After calculating the radiance for each point, we obtain the reflectance map of the image. The parameter of the reflectance map might be an unknown. In this case, the albedo and illuminant can be computed by assuming a Lambertian model with local surface smoothness. We can then estimate local surface normals, which can be integrated to give local surface shape. The techniques can be divided into four groups: minimization approaches obtain the solution by minimizing the energy function [26], [27], [28], Propagation approaches propagate the shape information from a set of surface points to the whole image [29], [30], [31], Local approaches derive the shape based on the assumption of surface type [32], [33], and Linear

approaches compute the solution based on the linearization of the reflectance map [34], [35]. The output of shape from shading is the surface orientation or needle map.

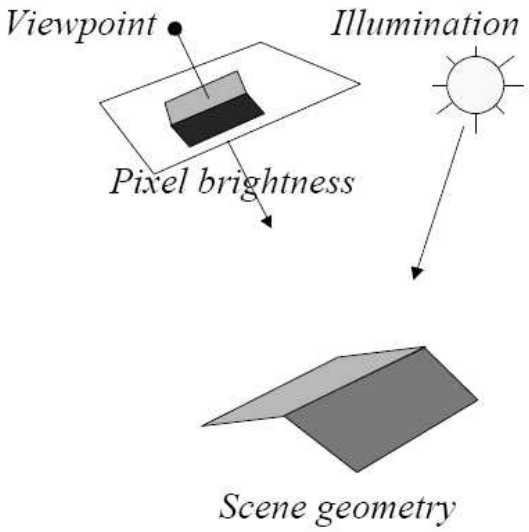


Figure 2.10: The principles of shape from shading

Shape from Motion (SFM). It refers to the process of extracting a 3D model of a scene from the spatial and temporal changes occurring in an image sequence or video of a moving rigid object (see figure 2.11). It exploits the relative motion between camera and scene. Regarding correspondence, the fact that motion sequences provide many closely sampled frames for analysis is an advantage. Firstly, tracking techniques, which exploit the past history of the motion to predict disparities in the next frame can be used. Secondly, the correspondence problem can also be cast as the problem of estimating the apparent motion of the image brightness pattern (optical flow). Two kinds of methods are commonly used to compute the correspondence. Differential methods use estimates of time derivatives and require therefore image sequences sampled closely. This method is computed at each image pixel and leads to dense measurements. Matching methods use Kalman filtering to match and track efficiently sparse image features over the time. This method is computed only at a subset of image points and produces sparse measurements. Various extensions of their approach have been proposed recently, including the incremental recovery of structure and motion [36], [37]. The factorization approach has been generalized to perspective case [38], to incorporate uncertainty [39], for independent

moving objects [40], articulated objects [41], for dynamic objects [42], [43] and with outliers and missing points [44]. The output of shape from motion is the polyhedron or point cloud.

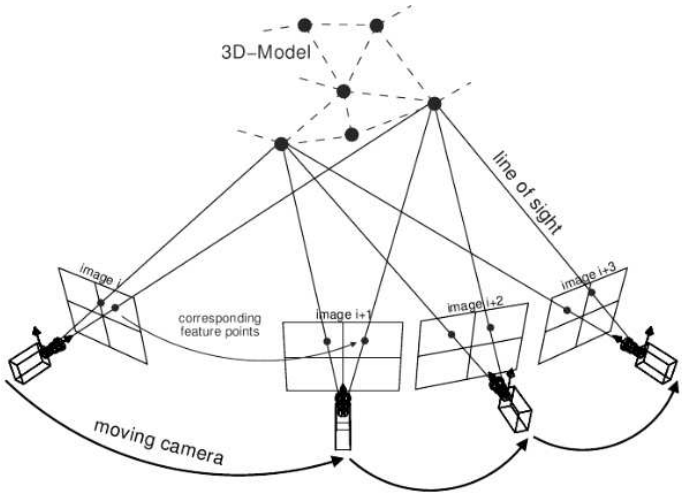


Figure 2.11: The principles of shape from motion

Shape from texture (SFT). It deals with the distortion of the individual texel (i.e., an element of the texture). As seen from figure 2.12, texture provides cues about shape or distance. Their variation across the image allows to estimate the shape of the observed surface. The shape reconstruction exploits the perspective distortion, which makes objects far from the camera appear smaller and denser. Also, foreshortening distortion makes objects not parallel to the image plane. The amount of both distortions can be measured (e.g. shape distortion and distortion gradient) from an image. To calculate the surface curvature at any point is a difficult problem. Therefore, the surface shape is reconstructed by calculating the surface orientation (surface normal). A map of surface normals specifies the surface's orientation only at the points where the normals are computed. But, assuming that the normals are dense enough and the surface is smooth, the map can be used to reconstruct the surface shape. Early work on shape from texture were feature based methods [45], [46]. An alternative approach uses spectral information [47], [48] like Fourier transform, wavelet decomposition and Gabor transform. Most of the techniques [49], [50] assumed that the textures were stationary. Forsyth [51] solves SFT

without making any assumptions of isotropic, homogeneity or stationary as was done by earlier works. The output of shape from texture is the disparity map or depth map.

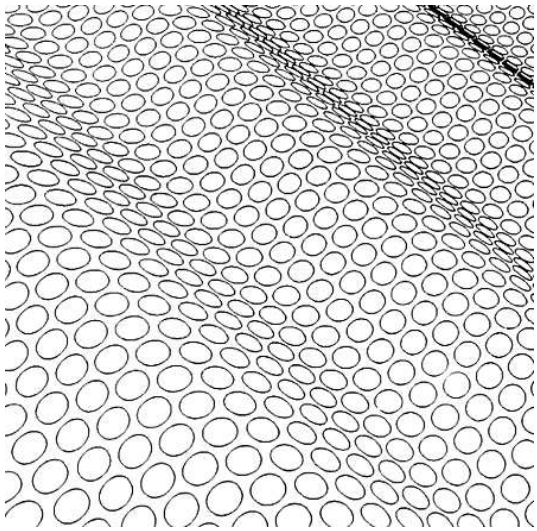


Figure 2.12: Example of texture distortion

2.2 Depth from focus and defocus (DFF/DFD)

Recently, another prominent technique, shape from focus/defocus has gained remarkable interest. Traditional DFF and DFD belong to imaging cue based category.

Depth from focus (DFF). According to image formation, only particular plane is brought into focus while, other parts of the image become blur. This blur can be used as the cue for depth estimation. The focus and the depth are also directly related. Depth from focus (DFF) applies this technique by using the sharpness of the scene to indicate how focused it is and recover the depth, respectively. The method requires several images to be taken with small incrementing focus settings [52], [53],[54]. The aim is to search for the best focused point (with minimum or zero blur) through the image stack. By using camera parameters and thin lens equation, the depth of this particular image point can be retrieved. The focus measure at each pixel location is the critical calculation used to define sharpness level around a pixel. Numerous criterion have been proposed in the

literature including gradient, Laplacian, and energy of the texture. Some related work done in DFF field are summarized in table 2.1.

	Summary	Reference
Choi *et al.*	recovered 3D shape from focused image surface instead of planar image frames. By locating the mode of the criterion function, it would be equivalent to the getting the sharpest focus.	[55]
Darrel *et al.*	generated Laplacian and Gaussian pyramids to calculate the sharpness map of scene, using the pipelined image processing hardware.	[56]
Horn	described a Fourier-transform method in which the normalized high-frequency energy from a 1-D FFT is used as an objective criterion.	[57]
Tenenbaum	used a thresholded gradient magnitude in which the Sobel operators are used to estimate the gradient. The criterion function is the sum of gradient energy over a local window centered around all pixel locations in the image.	[58]
Krotkov	applied a Fibonacci search over the focus measure profile for any pixel location across the set of multi focus images.	[57]
Jarvis	introduced sharpness measures based on entropy, variance and gradient.	[59]
Nayar *et al.*	developed a sum-modified-Laplacian operator to obtain local measures of the image focus quality.	[53]
Subbarao *et al.*	proposed the energy maximization of unfiltered,low-pass filtered, high-pass filtered and band-pass filtered images as focus measure functions.	[60]
Shafer *et al.*	proposed a combination of Fibonacci search and curve fitting for finding the minima of the Tenegrad focus criteria based error profile. Due to the window based approach and local equifocal assumption in conventional focus measure, these techniques result in artifacts near the edges in the image.	[61]
Ning *et al.*	proposed a focus measure handling artifacts at these edges, and apply graph cuts to obtain smooth omnifocus image.	[62]

Table 2.1: Summary of DFF related works

Depth from defocus (DFD). Instead of finding out the best focus point in DFF, method of DFD measures the blur amount (e.g. defocus parameters) explicitly, and infer the geometry of the scene. The method requires as few as two images with different optical geometric settings to evaluate the difference of blur level between each point in defocused images[63], [64], [61], [65], [66]. Pentland introduced one of the earliest approach to recover depth by defocused image. He simply used the pin hole model to produce sharp and large aperture image. As the distance between the imaged point and surface of exact focus changes (either increasing or decreasing), the imaged objects become progressively more defocused. The estimate change in focus is then used to relate the depth of the scene. Several models have been studied, and assume to define the blurring characteristic. The Gaussian model, the wave optic model, and the geometric optics model are the examples of Point Spread Function (PSF) assumption. Some related works in DFD are summarized in table table 2.2.

	Summary	Reference
Markov *et al.*	modeled shape and appearance of defocus as a Markov random field. But, the computational cost of theses techniques is quite high.	[67]
Surya *et al.*	used a spatial domain convolution/deconvolution approach. With this method, they modeled the image with a local cubic polynomial and the PSF as an arbitrary circularly symmetric function.	[68]
Watanabe *et al.*	proposed a Fourier domain approach to utilize the defocused information, and obtain the depth information of an object.	[69]
Nayar *et al.*	projected an active illumination pattern on the scene. A defocused set of the images is captured, which is then used for shape estimation.	[6]
Gokstorp *et al.*	obtained a multi resolution local frequency representation of the input defocus image pair. It is used to estimates the blur in the scene and geometry of the scene.	[70]

Mennucci *et al.*	studied and analyzed the accommodation cues for shape reconstruction. Accommodation cues are defined as all measurable properties of the images, which are associated with a change in the geometry of the imaging device.	[71]
Soatto *et al.*	proposed a solution to the generic bilinear calibration and estimation problem for 3D scene reconstruction.	[72]
Favaro *et al.*	introduced an optimal method to infer 3D geometry from defocused images. It involves computing orthogonal operators, which are regularized via functional singular value decomposition.	[73]
Favaro *et al.*	explained depth recovery approach via diffusion process	[74]
Zhou *et al.*	adapted new acquisition system by using coded aperture pairs for depth from defocus	[75]
Moreno *et al.*	worked on a new DFD approach based on active refocus of image and videos	[76]
Hermans *et al.*	recovered depth by sliding projector (velocity related)	[77]
Zhou *et al.*	proposed a very new challenging of depth retrieval by using only single image	[78]

Table 2.2: Summary of DFD related works

In comparison, the DFD has advantage over the DFF during image acquisition process, when scene objects may change their position dynamically. Overall, it requires less number of images, and therefore less computation. However, it needs more information about the defocused characteristics of the camera and generally returns less accurate result than in the DFF. The DFF/DFD is an example of the case where it can be specified either passive or active approach depending upon whether or not it is possible to project a structure of light onto the scene. Further substantial works of the DFF and DFD are detailed in the next chapter.

2.3 Summary of literature

Although a lot of solutions for the depth recovery problem have been introduced, still the goal of developing real time, robustness, and efficient algorithm for all kind of scenes has not been achieved. We have presented the key methods for 3D reconstruction in the field of computer vision. Each method has strength and weakness regarding the type of object and the system conditions. In the next chapter, we summarized the useful theoretical background for depth recovery system. Furthermore, some important works related to the DFF and DFD approaches are also explained.

Chapter 3

Depth from focus and depth from defocus

In this chapter, we first investigate the theoretical background of the methods used in this dissertation, such as focus and defocus geometric optics, camera parameters, Lens law, point spread function and blur analysis. The related selection of DFF and DFD works are also discussed in the second part.

3.1 Theoretical background

3.1.1 Camera geometry

The simplest image formation of a 3D scene is often explained by a pinhole model. As seen from figure 3.1, the ray of light passes through a pinhole and forms an inverted image of the object on the image plane. If the pinhole is narrow down (small aperture), only small amounts of light can get in. In this case, the image sharpness is limited by the diffraction. Through the small size of aperture, the light is scattered in many direction. Therefore, in practice, the lens is used to duplicate the pinhole geometry with similar focal distance, but larger aperture to admit more light.

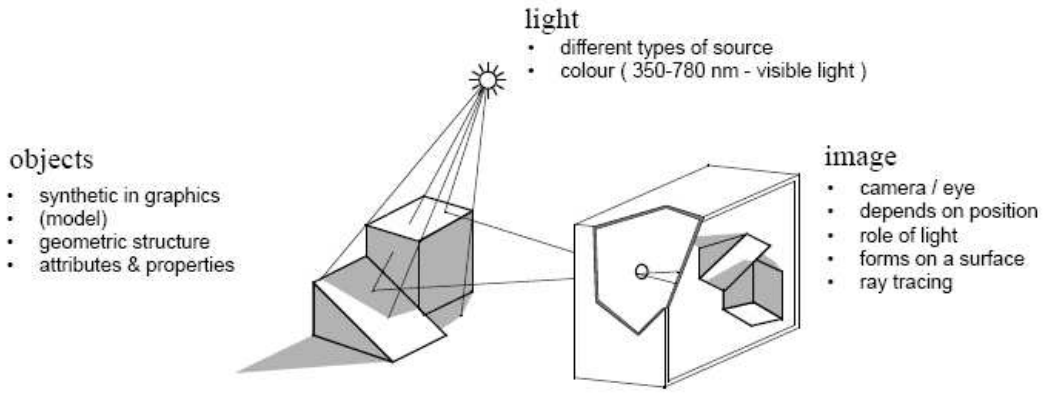


Figure 3.1: Pinhole Model

The scene can be captured either sharp or blur, depending on the convergence of all light rays from a single point on the object. The object appears sharp, if each point on the object plane is projected onto the image plane. The simple lens imaging system (in focus) is illustrated in figure 3.2, where object is placed at a distance u along the axis in front of a positive lens of focal length f and aperture diameter D. A sensor is placed at a distance v behind the lens. For simplicity, the thickness of lens in air is negligible. The distance from the object to the lens and the distance from the lens to the image are given by the thin lens equation:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \tag{3.1}$$

where

$$\begin{cases} f & \text{is the focal length.} \\ u & \text{is the object distance.} \\ v & \text{is the image distance.} \end{cases}$$

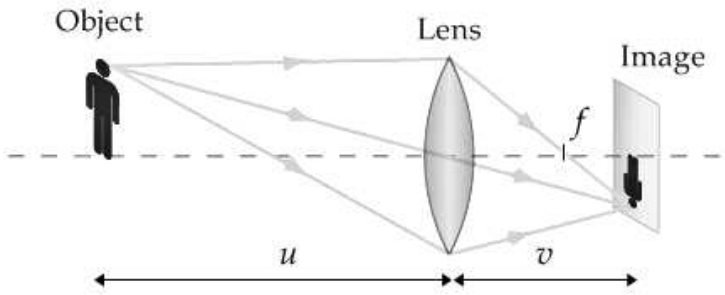


Figure 3.2: Imaging system in focus

3.1.2 Circle of Confusion

In the defocus case, when some of the sensor planes and the image plane are misaligned (see figure 3.3), the images are distributed over a circular patch called Circle of Confusion (CoC) on the sensing element, resulting in blur image [5]. The blur level can be determined from the diameter of CoC, which also increases in proportional to the distance from the object in focus. This phenomenon allows us to estimate the geometry of the scene by measuring the amount of blur in the image.

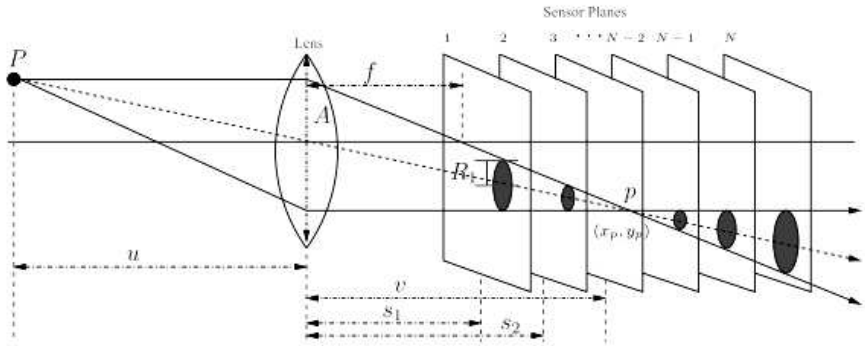


Figure 3.3: Circle of confusion from multi-focus system

Figure 3.4 represents the camera geometry of the single convex lens, with variable camera parameters (s , f , D).

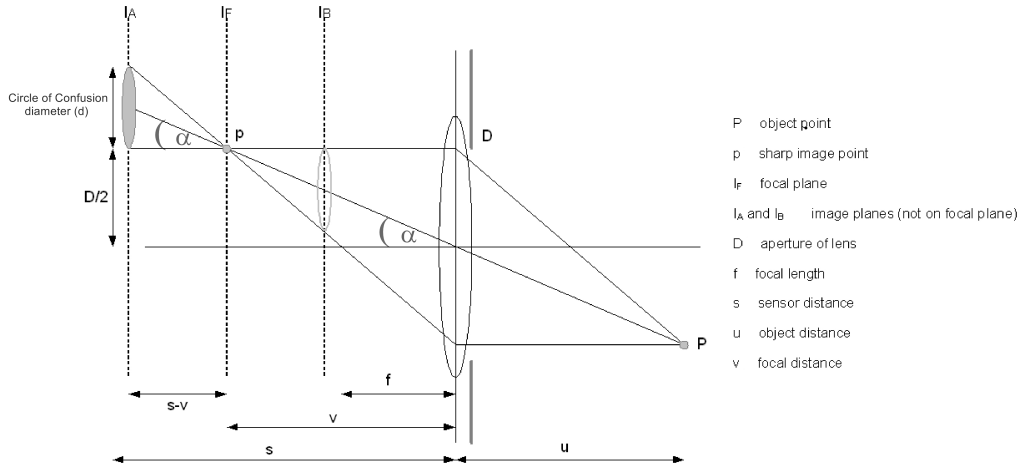


Figure 3.4: Camera geometry of single convex lens

Here, s specifies the sensor distance, f the focal length, and D the aperture diameter. The camera parameter setting is denoted by e_i where $e_i = (\ s_i\ , f_i, D_i\ )$ is a vector representing the i^{th} camera parameter setting. According to paraxial geometric optics, in order to define the radius of the circle of confusion, the knowledge of similar triangles is applied

$$tan\ \alpha = \frac{D/2}{v} = \frac{d/2}{s - v} \tag{3.2}$$

By employing the Lens law (Eq. 3.1), the normalized diameter and the radius of CoC becomes

$$d = Ds\Big(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\Big) \tag{3.3}$$

$$R = \frac{Ds}{2}\Big(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\Big). \tag{3.4}$$

3.1.3 Model of defocus

In this section, we investigate the model of defocus in spatial domain analysis, and Fourier domain analysis.

Spatial Domain Analysis

An image $I(x)$ captured on the image sensor is formed a sharp image $I_0(x)$, which can be blurred by the optics, the defocus, and the finite size of the pixels. We can describe the image $I(x)$ via the convolution of in-focus image $I_0(x)$ with the point spread function $h(x)$ as follow:

$$I(x) = h(x) * I_0(x) + e(x) \tag{3.5}$$

where

$$\begin{cases} x = \{x, y\} & \text{is an image coordinate in the two dimensional image, and} \\ e(x) & \text{is an additive noise.} \end{cases}$$

The point spread function $h(x)$ can be considered as a convolution of several blur kernels as shown in the figure 3.5.

$$h(x) = \int\int \rho(x - I_0)\eta(I_0 - t)\mu(t)dI_0dt \tag{3.6}$$

where

$$\begin{cases} \mu(x) & \text{is the optical blur,} \\ \eta(x) & \text{is the defocus blur, and} \\ \rho(x) & \text{is the sampling blur.} \end{cases}$$

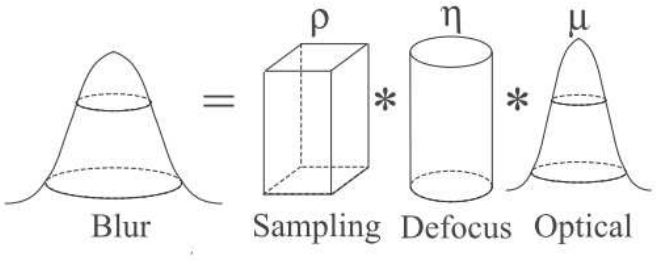


Figure 3.5: The overall blur kernels

According to [79], accurate models of blur kernel ($\mu(x)$, $\eta(x)$, and $\rho(x)$) are technically difficult to determine. It requires a special software and detailed physical model of the lens and the sensor, which are generally unavailable. Instead, we can consider one well-known measure of lens quality called Modulation Transfer Function (MTF). It is defined as the ratio of relative image contrast divided by the relative object contrast, where the object is the sinusoidally varying brightness at some spatial frequency [80]. The MTF depends on the location on the image plane, and inversely proportional to the spatial frequency. Without the relationship of MTF and the lens resolution, the optical blur kernel $\mu(x)$ can not be derived. Regarding the F-number of the lens, the blur decreases at lower F-number due to the improvement of the optical wavefront, and then increases at higher F-number due to the increasing diffraction blur, until it reaches the diffraction limit [81]. With respect to figure 3.6, the optical kernel can be written as,

$$\mu(x) = \left[\frac{2J_1(\gamma)}{\gamma}\right]^2 \tag{3.7}$$

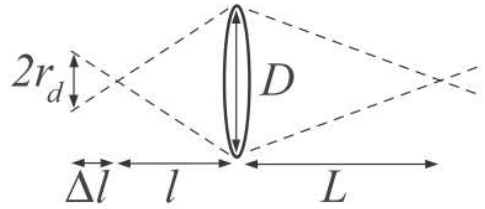


Figure 3.6: The defocus spot size

Here, J_1 is a Bessel function [82],$\gamma = \frac{\pi|x|D}{\lambda l}$, $\lambda = 0.7\mu m$ is the wavelength of the light, l is the distance from the lens to the image plane, and D is the lens diameter (see figure 3.6). In practical cases $l \approx f$, where f is the focal length of the lens. Using the definition of the F-number, $F \equiv \frac{f}{D}$, we can write $\gamma \approx \frac{\pi|x|}{\lambda F}$.

For the second blur kernel, the defocus kernel is assumed to have a cylindrical shape given by

$$\eta(x) = \begin{cases} \frac{1}{\pi r_d^2}, & |x| \leq r_d \\ 0, & |x| > r_d \end{cases} \tag{3.8}$$

Here, r_d^2 is the radius of the cylinder [83]:

$$r_d^2 = \left|\frac{\Delta l}{l}\frac{D}{2}\right| = \left|\frac{\partial l}{\partial L}\frac{\Delta L}{l}\frac{D}{2}\right| = \left|\frac{f^2}{L^2}\frac{\Delta L D}{l}\right| \approx \left|\frac{f^2}{L^2}\frac{\Delta L}{F}\right|. \tag{3.9}$$

The derivative $\frac{\partial l}{\partial L}$ is obtained from the thin lens formula, connecting between the focal length f, the distance L to the object, and the distance l to the focused image of the object (see figure 3.6)

$$\frac{1}{f} = \frac{1}{l} + \frac{1}{L}. \tag{3.10}$$

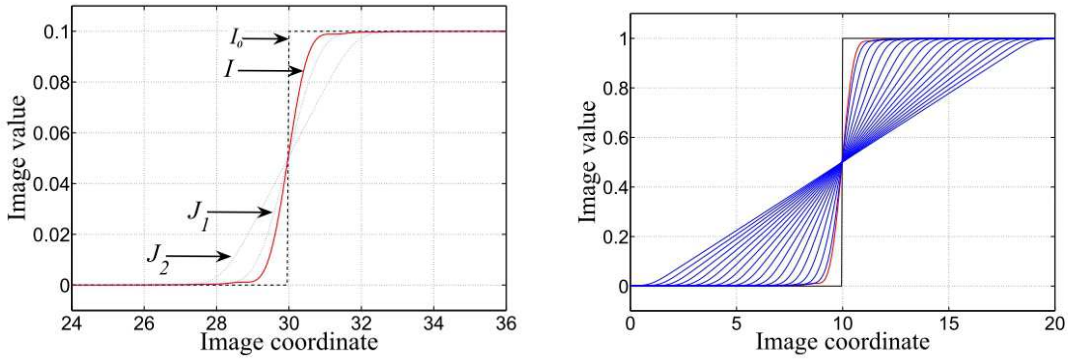
Lastly, the sampling kernel $\rho(x)$ describes the averaging over a square pixel of size Δx

$$\rho(x) = \begin{cases} \frac{1}{\Delta x^2}, & \max(|x_1|, |x_2|) \leq \frac{\Delta x}{2} \\ 0, & \text{otherwise} \end{cases} \tag{3.11}$$

Let assume an image in focus $I(x) = \rho * \mu * I_0(x)$ and a defocused image $J(x) = \rho * \eta * \mu * I_0(x)$. The minimal depth resolution ΔL is the depth change in which the difference between these two images becomes distinguishable.

The simulation result of the diffraction and defocus blurs are represented in figure 3.7. The sharp pre-image $I_0(x)$ has a step-like profile as shown in black dashed line. The focused image of this profile blurred only by the optics diffraction with $\lambda F = 1$, is shown by a red solid line; $I = \mu(1) * I_0(x)$. The blue dotted lines show the defocused images, blurred by defocuses with $r_d = 1$ and $r_d = 2$; $J_1 = \eta_1 * \mu_1 * I_0$and $J_2 = \eta_2 * \mu_1 * I_0$. Arbitrary small defocus blur will result in some difference between the focused and defocused blurs. In practical cases such difference is obscured by the image quantization and the noise.

Therefore, the difference between the images should exceed the quantization and noise level, in order to be reliably detectable.



(a) (b)

Figure 3.7: Step-like profile on sharp pre-image, smoothed by optics blur and different defocus blurs.

Fourier Domain Analysis

In Fourier domain, the blur can be depicted by a low-pass filter, which suppresses high frequencies. The cutoff starts from some characteristic frequency $\upsilon \approx \frac{1}{\varsigma}$, where ς is the scale of the blur kernel in the spatial domain. At the reasonable signal to noise ratios, the scales of defocus and optical blurs must be comparable. This implies that in the Fourier domain, their characteristics frequencies must also be comparable. The diffraction blur in Fourier domain is transformed into

$$F(\mu(x)) = \begin{cases} F^2(\varsigma - sin\gamma), \ \sqrt{u^2 + v^2} \leq \frac{2}{\lambda F} \\ 0, \ \sqrt{u^2 + v^2} > \frac{2}{\lambda F} \end{cases} \tag{3.12}$$

where the scale of the blur kernel $\varsigma = 2arccos(\lambda F \frac{\sqrt{u^2+v^2}}{2})$, u and v denote spatial frequency parameters in the x and y direction, respectively.

The defocus blur then becomes

$$F(\eta(x)) = \frac{L^2 F}{2\pi f^2 \Delta L \sqrt{u^2+v^2}} J_1\left(\frac{2\pi f^2 \Delta L}{L^2 F}\sqrt{u^2+v^2}\right). \tag{3.13}$$

Finally, the sampling blur kernel is given by

$$F(\rho(x)) = \frac{1}{2\pi}\frac{1}{\Delta x^2} sinc\frac{u\Delta x}{2} sinc\frac{v\Delta x}{2}. \tag{3.14}$$

Let assume two images: in-focus image I and out of focus image J. Then, $I(x) = h(x, L) * I_0(x)$ and $J(x) = h(x, L + \Delta L) * I_0(x)$. In Fourier domain, we obtain

$$F(I) = F(\rho(x))F(\mu(x))F(I_0(x)). \tag{3.15}$$

$$F(J) = F(\rho(x))F(\eta(x))F(\mu(x))F(I_0(x)). \tag{3.16}$$

The minimum depth resolution ΔL_{min} is the value in which the difference between $F(I)$ and $F(J)$ is detectable.

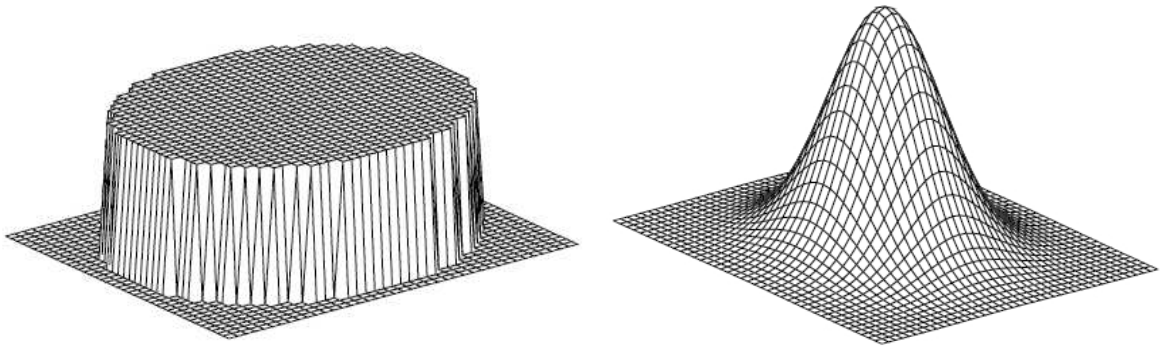
3.1.4 Point Spread Function

The PSF can be used to describe the imaging system response to a point input. It is analogous to the impulse response for a focused optical system. A point input can be either point source or point object representing as a single pixel in the ideal image. However, it will be reproduced by more pixels in the real image. The PSF in many contexts (e.g. defocus domain) can be thought of as the extended blob in an image. The degree of blurring (i.e., spreading) of the point object can determine an imaging system quality. The image of an object is therefore the convolution between the true object and the PSF.

$$I(x, y) = I_0(x, y) * h(x, y) \tag{3.17}$$

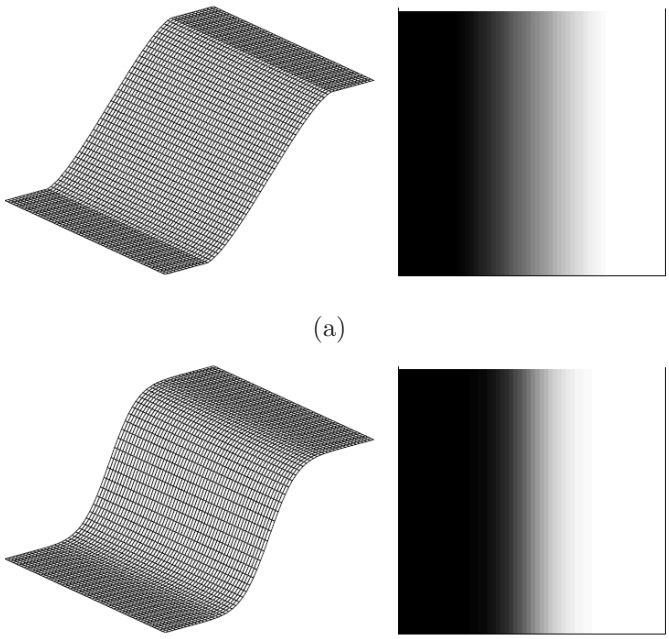
Several approximated models of the PSF have been investigated including the pillbox and Gaussian function. The representations and intensity plots of these 2 possible PSF

models are illustrated in figure 3.8 and figure 3.9. Each model of blur function had its own advantages and disadvantages.



(a) (b)

Figure 3.8: Point spread Function (a) Pillbox model, (b) Gaussian model



(a)

(b)

Figure 3.9: Intensity graph and corresponding images of (a) Pillbox model, (b) Gaussian model

The pillbox function can be seen as cone of light emerging from the lens with the point of the cone in the focal plane. If the sensor plane is shifted from the focal plane, then it

cuts the cone in circle with the diameter d. The pillbox function as the PSF model is

$$h_p(x,y) = \begin{cases} \frac{4}{\pi d^2}, \; x^2 + y^2 \leq \frac{d^2}{4} \\ 0, \text{ otherwise} \end{cases} \tag{3.18}$$

However, if within the circle, the brightness is not uniform, the PSF of the camera system at a point (x, y) may approximately modeled as a circularly symmetric 2-D Gaussian function. In addition, for the conceptual simplicity, the Gaussian model becomes more practical usable reference for PSF model when paraxial geometric optic is used and diffraction effects are negligible. The Gaussian function as the PSF model is

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{\left(\frac{-x^2+y^2}{2\sigma^2}\right)}. \tag{3.19}$$

If the brightness is assumed to be constant over a region of the image projected onto the CCD element, the spread parameter σ is proportional to the diameter d.

$$\sigma = kd, \tag{3.20}$$

where k is a constant of proportionality characteristic for every camera and can be determined from the camera calibration.

Figure 3.10 illustrates the original light pattern image compared to the resultant images after Gaussian blur and lens blur. The output suggests that Gaussian blur and lens blur provide similar effect. Another advantage is that the Fourier transform of a Gaussian function is also a Gaussian.

The Fourier transform of the PSF is called the Optical transfer function (OTF):

$$\hat{h}(\omega_x, \omega_y) = e^{-\frac{\sigma^2(\omega_x^2+\omega_y^2)}{2}}. \tag{3.21}$$

Thus, the defocus image in Fourier domain becomes,

$$\hat{I}(\omega_x, \omega_y) = \hat{I}_0(\omega_x, \omega_y) e^{-\frac{\sigma^2(\omega_x^2+\omega_y^2)}{2}}. \tag{3.22}$$

Eventually, once we know the input and output images, we can determine the PSF.

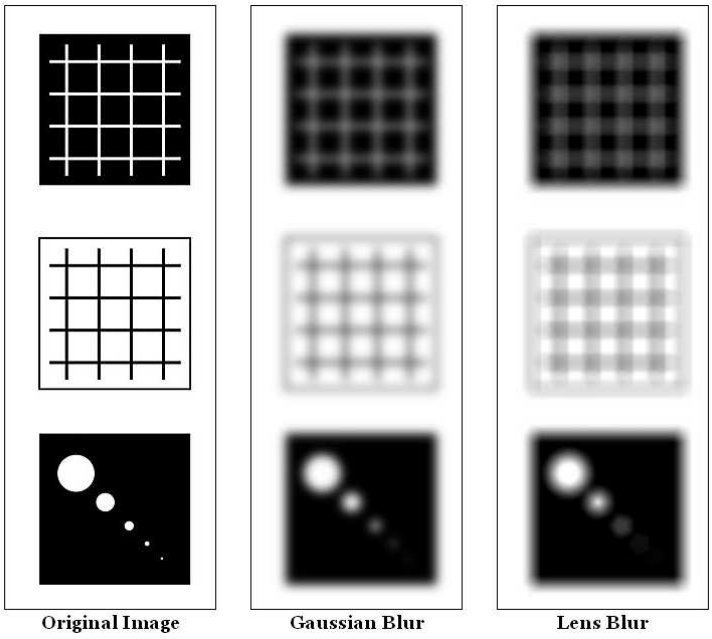


Figure 3.10: Comparison between Gaussian blur and lens blur

In the following section, we provide related works done using focus and defocus cues for shape estimation.

3.2 Depth from focus (DFF)

There has been considerable researches done on the the methods to recover the depth information from focus. Some of the main works proposed by Nayar *et al.* [53] and Subbarao *et al.* [84]. The method suggested that the different focus levels can be obtained by adjusting the camera parameters such as the lens to image plane distance v, the focal length f, or the aperture size D (see figure 3.4). Many observations for the different

camera parameters are done in order to estimate the focus measure using various criterion functions.

Choi *et al.* recovered the 3D shape from focus image surface instead of planar image frames [55]. The method is based on approximation of the Focused Image Surface (FIS) by a piecewise curved surface which tracks the realistic FIS in image space. The focused image surface in image sequences is illustrated in figure 3.11. The piecewise curved surface is estimated by interpolation using the Lagrangian polynomial. The methods involve in obtaining many images for the various camera parameters, and estimating the focus measure using various criterion functions.

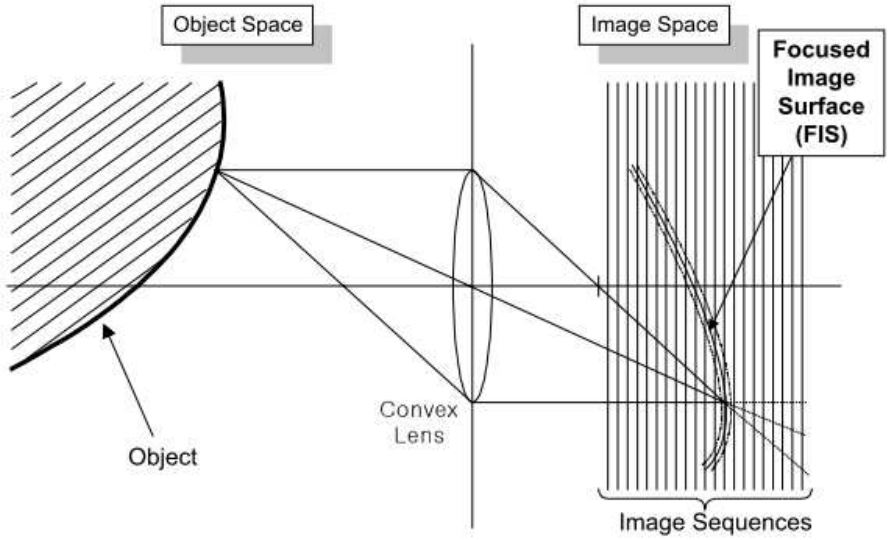


Figure 3.11: Focused image surface in image sequences [55]

Krotkov has experimentally evaluated several such criterion functions including the Laplacian and Teningrad operators [57]. Instead of using window-based blur estimation operators, Asada proposed a method in which the blur can evaluate from the intensity change along corresponding pixels in the multi-focus images [85]. The fundamental drawbacks of the DFF method is that it is time consuming for both image acquisition and depth estimation. Practically, about ten or more images are required to estimate the depth of a scene for a reasonable level of accuracy [86].

3.3 Depth from focus (DFD)

There are a number of papers in the literature addressing the methods to recover the depth information from defocus. The fundamental methodology is to estimate the depth from the measurement of the relative defocus between images. It can be classified into 2 types: passive DFD and active DFD.

Passive DFD. Two or more images of a scene are taken with different camera parameters. Based on the difference in defocus blur between at least two observations, the depth in the scene is estimated. This method was first introduced by Pentland [87], where he used a half mirror to split the observation and then passing the observation through two lenses with varying apertures. For each image, we can use a patch based Fourier transform and the ratio of the windowed Fourier transforms to obtain the relative defocus. If one of the images is in-focused, we can achieve the depth estimate from this relative blur.

Based on this work, Subbarao *et al.* considered the method of blur estimation from blurred edges [88]. They proposed a closed form solution for estimating the blur parameters and extended the earlier assumption of the point spread function (PSF) from being Gaussian to the more general case of being rotationally symmetric. In the later work, he also considered the usual case of depth estimation from defocus by allowing changing of more than one camera parameters simultaneously. The methods based on estimation of blur from step edges were too restrictive and could not be widely used. The approach for depth from defocus based on using two or more images from different camera parameters was further explored. In the earlier cases, ratio of windowed Fourier transforms were taken for estimating the depth in the scene. This approach was refined by Gokstorp [70]. In their work, a local frequency representation was adopted in which the local estimates of instantaneous frequency, amplitude and phase were computed. A set of complex-valued Gabor filters are used to obtain the representation. The final estimate of blur is computed by averaging the result from the various filters. A similar method was also proposed by Xiong *et al.* [61]. They proposed moment filters to address the problem of tuning the individual frequency components. The property of moment filters is that it is a polynomial

approximation with the order of the polynomial being tuned for obtaining a large number of narrow band filters. Due to the recursiveness in frequency and spatial domain moment, the filters can be implemented efficiently. While, the methods based on narrow band filters are attractive, it is typically computationally extensive and also the accuracy is a function of averaging over a large number of such filters. A different approach was adopted by Watanabe *et al.* [69]. They suggested to use the broadband rational filters. The technique uses a normalized ratio of near and far focused images, which is then convolved with a small number of broadband rational filters (invariant to the texture). However, a drawback of the method is under the assumption that the normalized ratio of the near and far focused images is linear in nature. This assumption is therefore valid only when the amount of defocus between the near and far images is very small, and will not work for a general class of defocused images. These concluded the passive DFD based on Fourier domain, the method based on spatial domain is summarized as the following.

Ens *et al.* suggested a method for estimating depth from defocus by using a matrix based approach [52]. They propose the estimation of the transfer function from the near image to the far image using a regularization based approach. An approach for depth from defocus using spatial deconvolution was proposed by Surya *et al.* [64]. This is achieved by using an operation termed S-Transform. By assuming a cubic polynomial form of an image, the convolution and deconvolution can be expressed in terms of a simple Laplacian operator modulated by the second moment of the PSF. The resultant depth maps obtained were however not very accurate.

Recently, Favaro *et al.* solve the problem of depth from defocus by considering two specific cases. The first case is when the form of the PSF is known, and the second case is when the form of the PSF is unknown. When the form of the PSF is known, they use a least squares solution in solving the projection between finite and infinite dimensional Hilbert spaces and this is achieved by estimating a set of orthogonal operators. They use functional singular value decomposition for estimating the operators. The values are truncated beyond a certain singular value and this results in regularization being incorporated into the solution. However, if the form of the PSF is unknown, they first

learn the set of projection operators from blurred images. Then, they use this learnt set of projection operators to estimate the shape in the least square sense as done previously.

Active DFD. Pentland introduced the active depth from defocus, where a known pattern of light is projected onto the scene [89]. The optical path of the projected pattern is maintained and similar to the observing camera by using an optical beam splitter. The structured light source projects a pattern on the scene while the camera records it. Since the original projected light source is known, the defocus blur introduced by the depth in the scene can be measured against the original pattern and the blur introduced due to defocus. Consequently, the depth in the scene can be measured. The method was refined by Noguchi *et al.* for microscopic shape from defocus [8].

Nayar *et al.* proposed a real-time focus range sensor [90], where the illumination pattern used for depth estimation is analyzed in detail. They performed a Fourier transform analysis in the various aspects of the defocus, illumination pattern, and the focus operator, in order to optimize the illumination pattern. Furthermore, the focus operator is tuned to enhance the depth estimation. These refinements were used to develop a robust real time focus range sensor. The comparison of tuned focus operators are illustrated in figure 3.12.

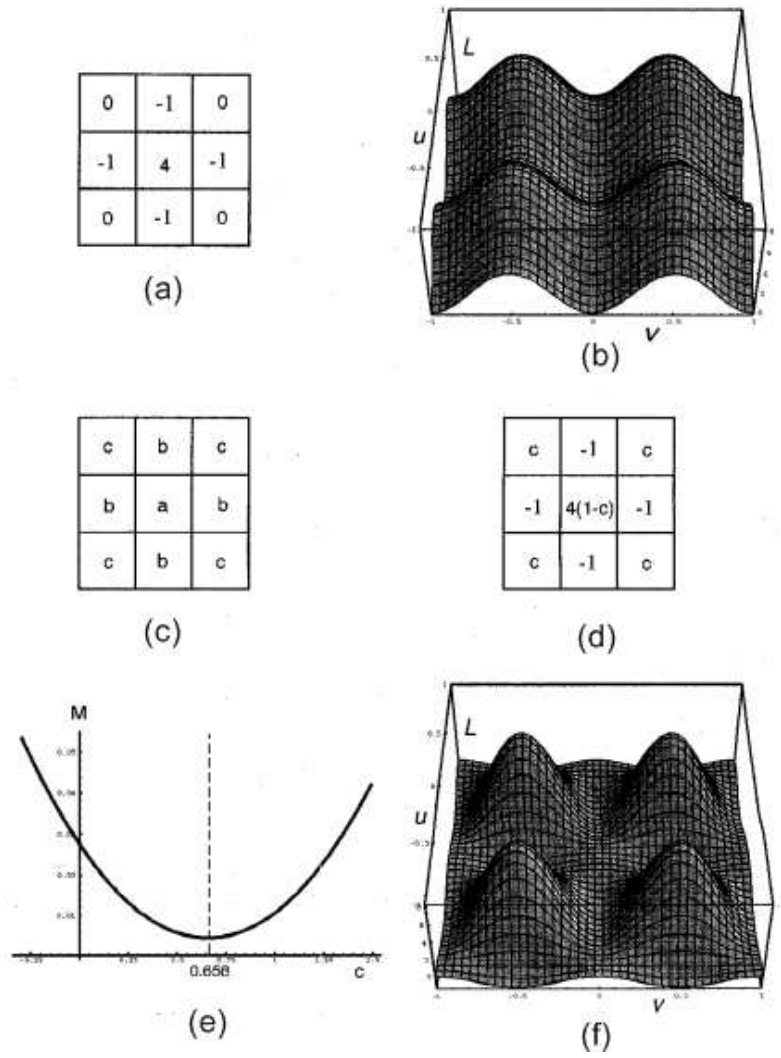


Figure 3.12: (a) The 3 x 3 Laplacian and its (b) Fourier transform, (c) The kernel structure for a 3x3 operator that is symmetric, (d) The kernel of a 3 x 3 operator that is insensitive to the zero frequency component, (e) The second moment of each of the four operator peaks is minimized, and (f) Response of the tuned focus operator (sharper peaks than the Laplacian) [90].

Moreno-Noguer *et al.* demonstrated active refocusing of images and videos [76]. A dot-pattern is used instead of a stripe pattern so that the dot-pattern projected can be easily removed while refocusing the image. They also use segmentation techniques for obtaining a dense depth map. A similar work has been done by Zhang *et al.* [91] where they use the projector for projecting a shifting light pattern and a set of images of the scene is taken. The depth of a pixel is computed by analyzing the temporal variation of the pixel

brightness due to the defocus. While, these works demonstrate the effectiveness of active depth from defocus, the main drawback is that they require elaborate set-up and can be done only in carefully controlled environments with calibrated illumination conditions. These requirements preclude the use of these techniques in outdoor environments and natural conditions.

3.4 Our approach

Our approach uses novel active range sensor and combines both depth from focus and depth from defocus with the help of light pattern projection. This method falls into both the imaging cues approach and the active structure light based approach. The aim is to introduce a new and alternative approach to solving some of the specific problems found in classical approaches, such as the weak texture surface and occlusion problem. With this approach, the projected light pattern images are acquired within certain ranges, similar to the DFF approach, but the numbers of captured images are much smaller, and the images do not need to be sharp. In traditional DFD, blur estimation is a very difficult problem because a point that represents the defocus information has contributions from several PSF that are induced by different depths. Our method avoids this problem because we can control the deformation by placing an additional semi-transparent screen after the light source. Therefore, by considering the light pattern as a plane, one point on the object representing defocus information corresponds to only one PSF. Moreover, when the projected patterns are out of focus, we assume that their PSFs follow a Gaussian distribution. Eventually, the relationship between the set of PSFs obtained from different blurs and the variation of the object depths can be determined.

3.5 Chapter summary

We have presented the theoretical background and the fundamental knowledge that is useful for the depth recovery system. The defocus model has been analyzed in both spatial

domain and Fourier domain. Furthermore, we have discussed also about the Point spread function, which is one of the key parameter to our final depth derivation. Some previous interesting works on the DFF and DFD are also summarized. The detail explanations of our approach are presented in the following chapters.

Chapter 4

Methodology

In this chapter, we first discuss about our prototype principle including the light pattern projection system, the main optical system, and the acquisition system. With this solid knowledge, we then derive our depth estimation system. Lastly, we mention about the design and deformation of the dynamic light pattern which is another important part of the system.

4.1 Our prototype principle

A new prototype of range sensors has been developed. We integrate both depth from focus and depth from defocus with the use of dynamic structured light. A video projector is used as a light source, to produce strong projecting light patterns. It is much more powerful compared to normal lamps and much simpler compared to LEDs. Moreover, it moderates the additional pattern modifications that are needed for different types of surface textures.

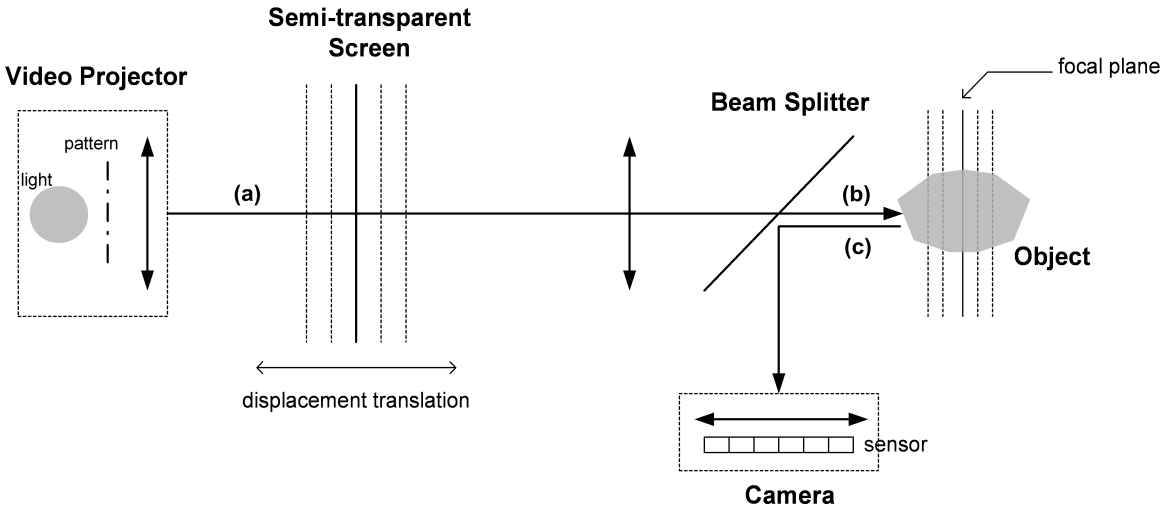


Figure 4.1: Proposed approach model

Figure 4.1 illustrates the overall design of the system. The main purpose of using a semi-transparent screen is to control the defocus level that corresponds to each screen position. It also helps to solve the magnification problem caused by the fact that the projected light patterns from the video projector are originally small compared to the patterns projected on the object without passing through the screen. Reducing the intensity of the powerful light is additional advantage. Because the normal lamp provides insufficient brightness, the video projector sometimes produces too strong a light, which can be adjusted by a projector setting or by putting in some blocking element(s). The beam splitter is mainly used to observe the object on the sensor, and it allows for the projection of a light pattern onto the object.

The entire set-up can be separated into 3 systems:

4.1.1 Light pattern projection system

In figure 4.2, the video projector projects the elementary light pattern of size $N \times M$ pixels. Consequently, we observe a light pattern of size $H \times L$ mm on the screen. The size of the input pattern from the video projector ($N \times M$) and the size of the pattern that appears on the screen ($H \times L$) indicate the resolution of the light pattern. We aim to

project the sharp elementary pattern of size P on the screen regardless of the positions. However, at some screen displacements, we may need to adjust the video projector to maintain the pattern sharpness.

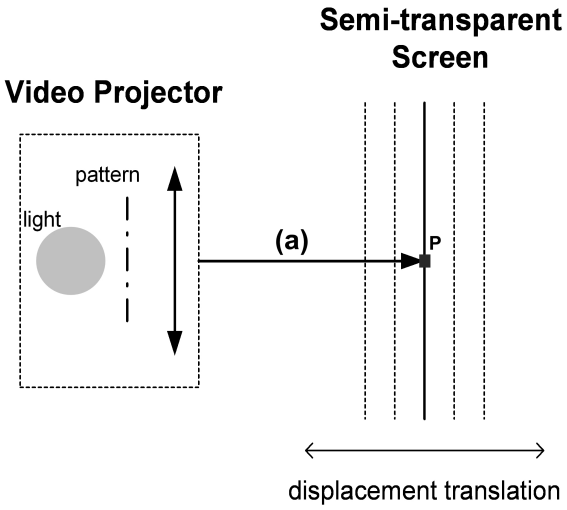


Figure 4.2: Light pattern projection system

4.1.2 Main optical system

The system is considered from the light pattern on the screen (size P), projecting through a specific optical system onto the object. The optical components of the system consists of a semi-transparent screen, a compound lens, and a beam splitter. An elementary size of the light pattern that appears on the object is denoted as P'. The optical path of the system is illustrated in figure 4.3.

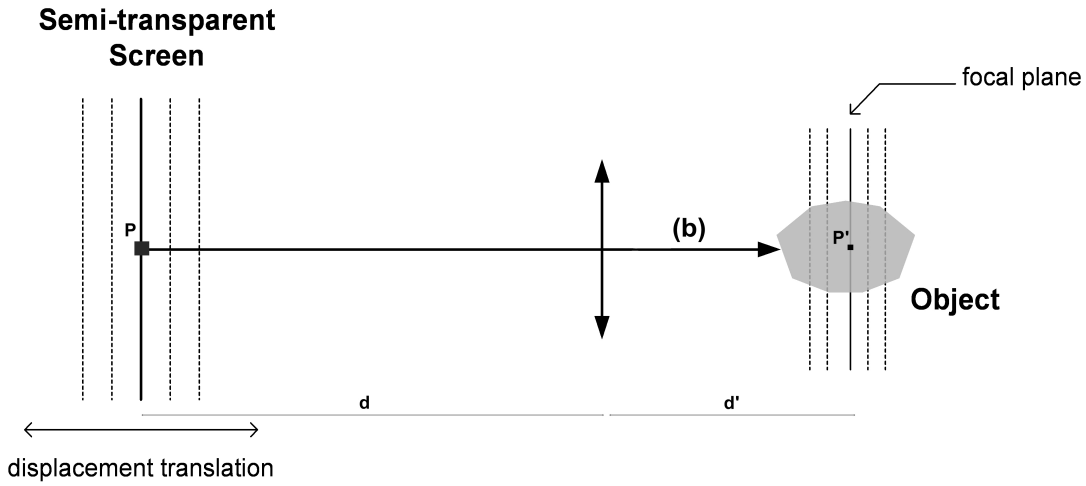


Figure 4.3: Main optical system

The magnification of this system can be written as

$$\gamma = \frac{P'}{P} = \frac{d'}{d}. \tag{4.1}$$

Thus,

$$P' = \frac{Pd'}{d}, \tag{4.2}$$

where

$$\begin{cases} \gamma & \text{is the main optical system magnification,} \\ d & \text{is the distance from the light pattern on the screen to the additional compound lens, and} \\ d' & \text{is the distance from the additional compound lens to the object.} \end{cases}$$

In general, for the ideal case where the object is placed in or very close to the surface of the best focus, an output image formed on the sensor is sharp or identical to the input. The relationship between the input and output images is the following:

$$I = I_0 * \delta \xrightarrow{\mathcal{F}} \hat{I} = \hat{I}_0. \tag{4.3}$$

However, our concern is deformation when a semi-transparent screen's displacements are varied at different depths of field. The blurring function has an influence on the system and

therefore needs to be taken into account. The defocused output image can be rewritten as the convolution between the input image and a blurring function h, as follows:

$$I = I_0 * h_{d,d'}, \tag{4.4}$$

where $h_{d,d'}$ are the blurring functions (PSF) corresponding to distance d and d'.

As explained in subsection 3.1.4 , we assume the PSF follows the Gaussian distribution. This extracted PSF will be used for depth computation.

4.1.3 Acquisition system

The acquisition system captures objects with projected light patterns (size P') to the sensor via a beam splitter. The observed pattern size on the sensor is denoted as P''. The optical path of the system is simplified, as shown in figure 4.4. The magnification of this system is

$$\gamma' = \frac{P''}{P'} = \frac{l'}{l}. \tag{4.5}$$

Hence,

$$P'' = \frac{P'l'}{l} = \frac{Pd'l'}{dl}, \tag{4.6}$$

where

γ' is the acquisition system magnification,
l is the distance from the light pattern on the object to the camera lens, and
l' is the distance from the camera lens to the sensor.

The relationship between the sizes (in pixels) of the pattern P'' on the sensor and of the elementary pattern from the video projector is used for the light pattern specification (width and density). Only one pixel of the light pattern (from the light source) may require several representing pixels on any of the captured images.

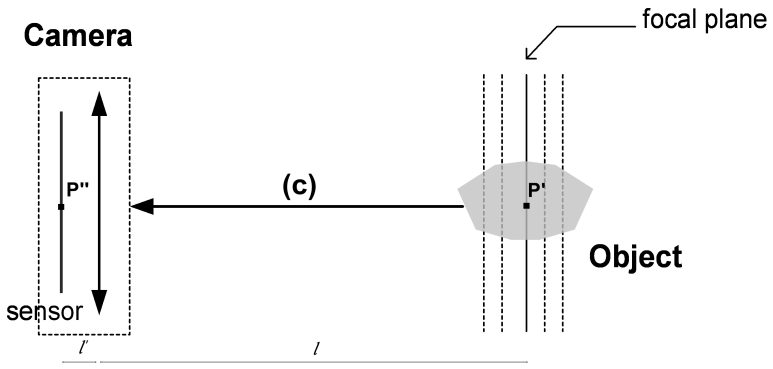


Figure 4.4: Acquisition system

This concludes the prototype principles. The corresponding practical implementation is detailed in Chapter 5. We now explain the theoretical relationship between the blur level (the spread parameter) and the depth.

4.1.4 Depth Estimation

We use the camera lens (compound lens) instead of single convex lens as shown in figure 4.5. Here, s specifies the sensor distance ($\left\|\overline{H'C'}\right\|$), f specifies the focal length ($\left\|\overline{H'F'}\right\|$), v is the object distance ($\left\|\overline{H'A'}\right\|$), and D is the aperture diameter. According to paraxial geometric optics, to define the radius of the circle of confusion, the knowledge of similar triangles is applied.

$$tan\ \alpha = \frac{D/2}{\left\|\overline{F'A'}\right\|} = \frac{\phi}{\left\|\overline{A'C'}\right\|} = \frac{\phi}{x}. \tag{4.7}$$

The normalized diameter of CoC can be rearranged into the following:

$$\phi = x \cdot \frac{D/2}{\left\|\overline{F'A'}\right\|}, \tag{4.8}$$

where

$$\begin{cases} x & \text{is the object distance and} \\ \phi & \text{is the diameter of Circle of Confusion (CoC).} \end{cases}$$

The higher the degree of blur is, the larger the CoC. The CoC is also proportional to the value of spread parameter σ and can be written as

$$\sigma = K_1 \phi, \tag{4.9}$$

where K_1 is a positive constant.

Given a spread parameter σ, aperture size D, and the controllable distance $\left\|\overline{F'A'}\right\|$, the only unknown of the system (see equation 4.8) is x. Eventually, we can derive the object distance x, which directly relates to the real depth of the object as

$$x = \pm \frac{2\sigma \cdot \left\|\overline{F'A'}\right\|}{K_1 D} \tag{4.10}$$

$$x = \pm K_2 \sigma \tag{4.11}$$

where K_2 is a positive constant for a given position of the screen.

The aim of the whole system is to achieve the object depth x. The extracted PSFs are used for spread parameter σ computation. However, K_2 is an unknown variable. Another concern is that the pattern on the screen at different displacements may not be constant, even after tuning a video projector for the best sharpness. Moreover, the distance l varies according to the object depths, which is also an unknown variable. To solve these problems, the calibration process is required. The depth is eventually computed by using a derived depth formula with the parameters obtained from calibration.

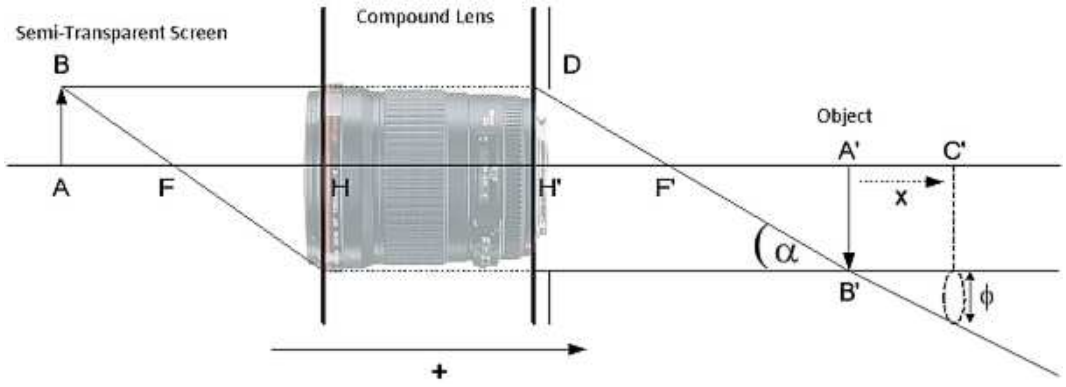


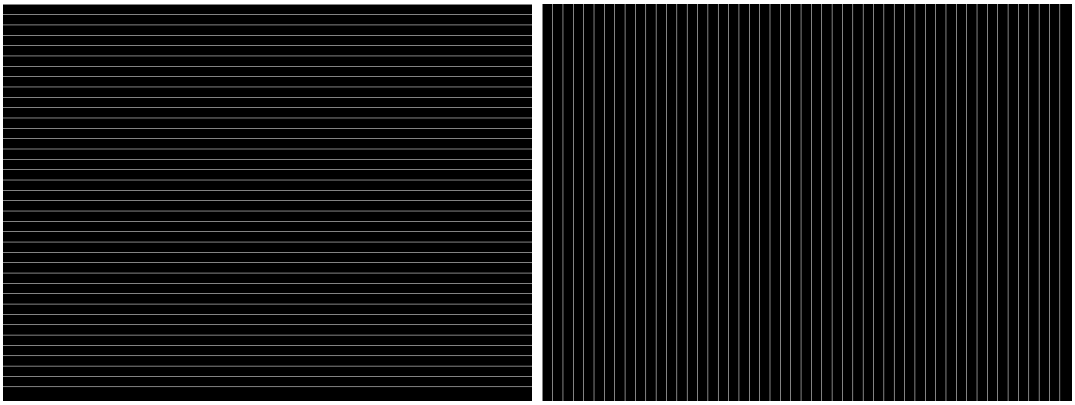
Figure 4.5: Camera geometry of compound lens.

4.2 Illumination pattern

4.2.1 Design of dynamic light pattern

All passive techniques share the same inherent weakness when the nature of an object's texture is poor. A weakly textured surface does not provide sufficient details for depth estimation because both the focus and defocus give the same representation. An effective solution to solve this problem is to employ an active illumination pattern. The structured light source projects a pattern on the scene through specific optical setting while the camera captures it. Because the original projected light source is known, the defocus blur introduced by the depth in the scene can be measured against the original pattern. The choice of an appropriate light pattern is important to optimize our final reconstruction. Highly textured light patterns are forced onto the object, improving the overall depth recovery system to be reliable and more precise. Moreover, to avoid rotational variance, it should be designed in a symmetrical or semi-symmetrical arrangement. The density of the projected pattern or its spatial frequency should correspond to the frequency of the height variation to be captured. For example, an object with a high level of detail requires a finer texture, whereas an object with smooth structural changes can use a sparse pattern instead. To be specific, for the object with small depth variations, we can reduce the processing time by projecting the sparse pattern instead. There are fewer intensity

profiles to be analyzed, and it returns sufficient results that are similar to the results from a denser pattern projection. For our experiment, we employ a set of parallel stripes with regular spacing. Spacing and shifting step sizes are determined from the scale of the texture to be analyzed. The width of the lines and the density of the pattern are selected according to equation 4.6, to cover as much of the reconstructed area as possible. Figure 4.6 illustrates a sample of a stripe light pattern with a width of 1 pixel and 20 pixels for spacing in between.



(a) (b)

Figure 4.6: Example of illumination pattern used (a) horizontal stripes (b) vertical stripes.

4.2.2 Deformation of projected texture

The projected texture is deformed by depth variation and physical properties of the surface. The deformation due to depth variation is the transformations of a ray of projected light due to variations in surface depth. This results in pattern shifting and pattern deformation. For pattern shifting, the position where a particular projected pattern is imaged by the camera depends on the absolute height from which the pattern is reflected. The amount of shift depends on the height difference between the objects as well as the angle between the projector axis and the height axis. While, for pattern deformation, any pattern that is projected on an uneven surface gets deformed in the captured image depending on the change in depth of the surface. These deformations depend on the absolute angle between the projector axis and the normal to the surface at a point as well as its derivative.

The deformations due to physical properties of the surface concern with other factors, such as reflectance and natural texture, that contribute to the changes in the projected texture. The physical property that mainly affects the final deformed pattern is the reflectance of the surface. Various problems due to the reflectance include specular surfaces, transparent surfaces, translucent surfaces, natural texture, and dark surface. The solution to these deformation problems is explained in Chapter 5.

4.3 Chapter summary

This chapter is dedicated to the prototype principle, which consists in 3 main systems: the light pattern projection system, the main optical system, and the acquisition system. The system components and their optical paths are explained in detailed. The second part of the chapter described the design of the light pattern and its textured deformation. In the next chapter, we present the detail explanations of our implementation.

Chapter 5

Implementation

This chapter is focused on the implementation of our new depth estimation system. In our experiment, a Canon SX80 Mark II video projector with a resolution of 1400x1050 is used as the light source. The horizontal stripe illumination patterns with a width of 1 pixel and 20 pixels for spacing are applied. The beam of the projecting light pattern then reaches the semi-transparent screen and an additional lens (Canon telephoto lens 135 mm). The light rays passing through the lens are split into 2 directions by the beam splitter. One beam is projected onto the object, and another is transmitted from the object to the sensor. The scene object is captured using a Canon EOS-1Ds Mark II camera with an attached 50 mm lens. The data flowchart illustrated in figure 5.1 describes the main operations. The process is divided into two steps: object calibration, and object reconstruction. The image acquisition, image profile analysis, pattern localization, and spread parameter calculation aim at extracting depths, and are performed for both calibrated objects and test objects. The object calibration process provides the reference map and the depth calibration. Depth calibration is performed only once and is also identical for the similar object material; while the object reconstruction process returns candidate depths in which the reference map is required to evaluate the final depth. The detailed implementation are described in the following sections.

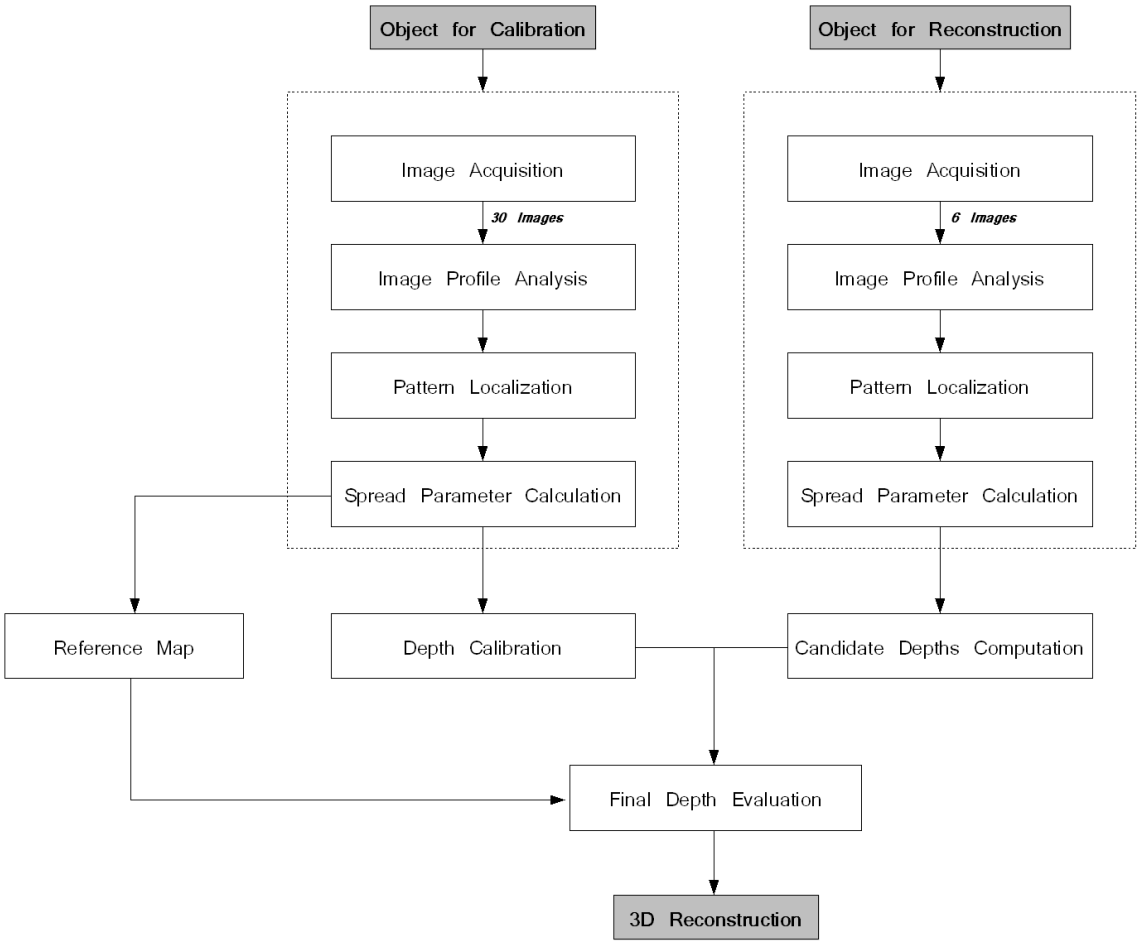


Figure 5.1: Flowchart

5.1 Image Acquisition

The camera settings (e.g., F, ISO, shutter speed) are carefully tuned such that the system keeps the whole object sharp in all of the images and only the defocused patterns experience varied deformation according to the object's depth. All of the optical components in this set-up are fixed. Only the semi-transparent screen is moved, which results in several scene images with different blur levels. With this specific setting, we can analyze the defocus of the light pattern projected on the object. The entire experimental set-up is shown in figure 5.2.

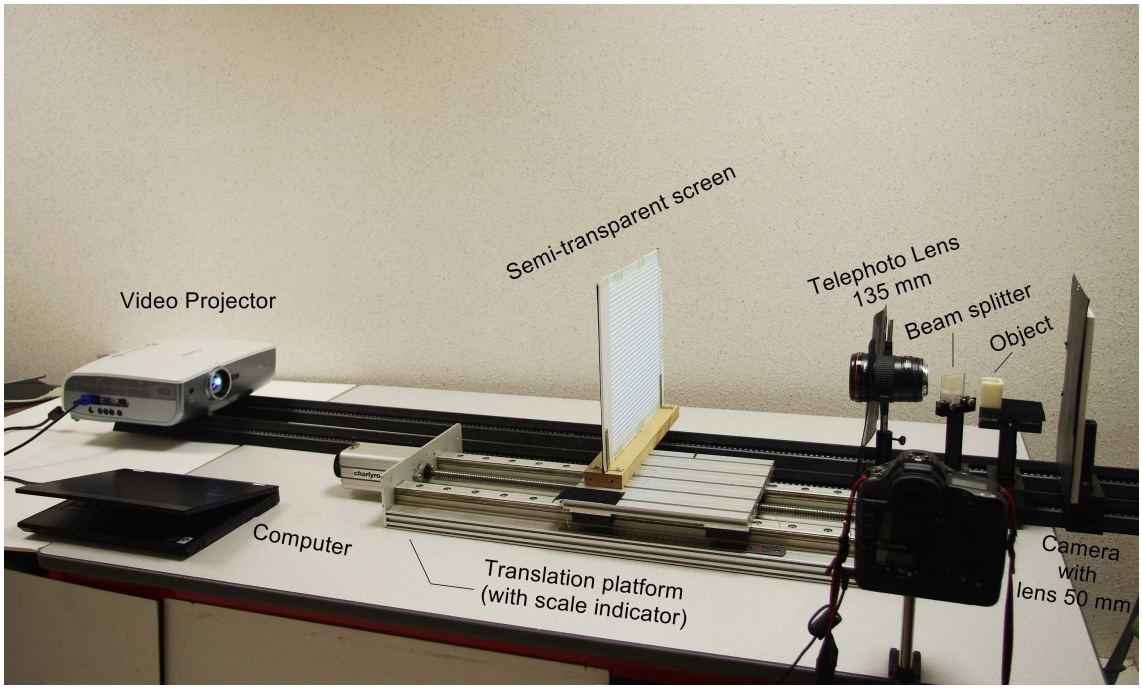


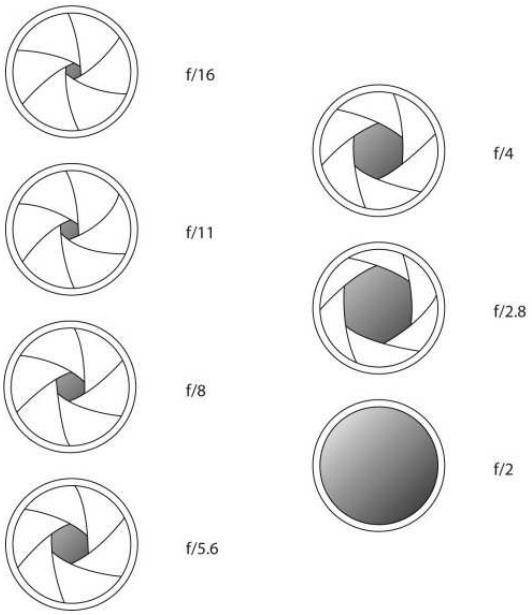
Figure 5.2: Experimental set-up

5.1.1 Camera setting

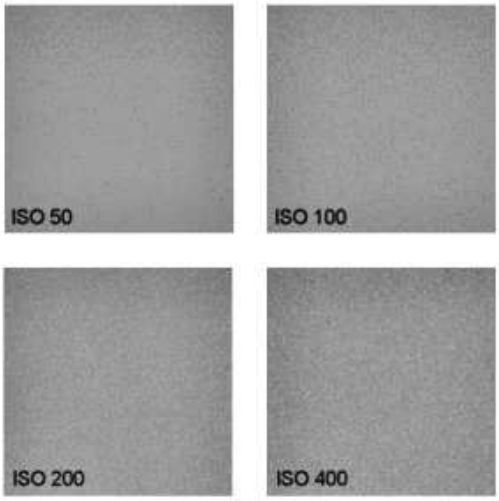
In order to determine the appropriate camera setting, we refer to the exposure triangle (see figure 5.3). It illustrates the relationship of the 3 important settings that directly affect the exposure, the quality, and the type of the captured images. The detail of the camera settings are summarized in table 5.1 and the corresponding impact of camera setting are in table 5.2.

Camera setting	Summary	Our setting
Exposure mode	By using the manual mode, we can specify the aperture, ISO and shutter speed regardless of whether or not these values lead to a correct exposure.	Manual(M)

Aperture	The function of the aperture is similar to the Iris in our eye. We can control the amount of light or how large the area where the light allow through the camera lens by adjusting the aperture. This is specified in terms of the F-stop value. The higher the number of F-stop, the smaller the aperture is. The smaller aperture also means the larger depth of field (DOF).	F/8



Shutter speed	It is used to control the amount of time, that the light is allowed to expose the image sensor on the camera. A slower shutter speed means the sensor is exposed to the light for a longer time and therefore it is suitable for less available light situation. However, the slower shutter speed also means a greater risk for motion blur (e.g. moving object, camera shake).	1.3 sec
ISO	The ISO speed is used to determine how sensitive the camera is to the incoming light. Similar to the shutter speed, it also correlates 1:1 with how much the exposure increases or decreases.	ISO 200



However, unlike the aperture and the shutter speed, the lower ISO speed is almost always desirable. Because the higher ISO speeds can dramatically increase the image noise.

Table 5.1: Camera setting

F-stop	Aperture size	Required shutter speed	Depth of field (DOF)
Higher	Smaller	Slower	Wider
Lower	Larger	Faster	Narrower

Table 5.2: Corresponding impact of camera setting

In our experiment, we work under the dark room with less available light from only the video projector source. Therefore, we choose a long shutter speed of 1.3 seconds to let the light exposed to the sensor for a long time. By employing this slower shutter speed, we also put the camera on the tripod to avoid the motion blur. While, for the aperture size, we set it at F/8 as it allows the light to the lens covering our working DOF range. Lastly, we select ISO 200 to minimize noise on the image.

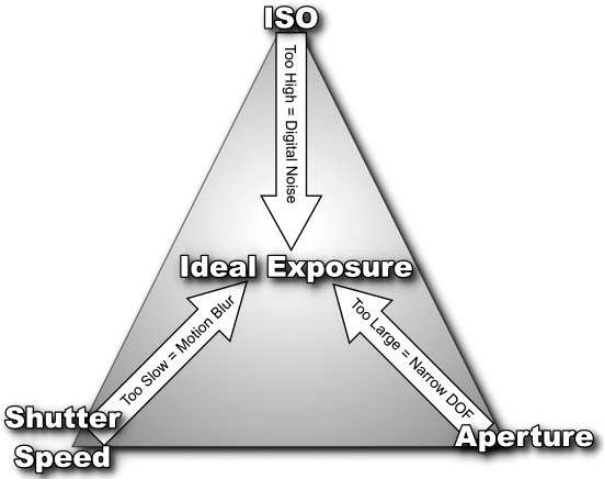


Figure 5.3: Exposure triangle

As explained in section 4.1, our prototype principle of the set-up consists in 3 main systems: the light pattern projection system, the main optical system, and the acquisition system. The practical set-up used in this experiment and its observation view are illustrated individually in figure 5.4, 5.5, and 5.6.

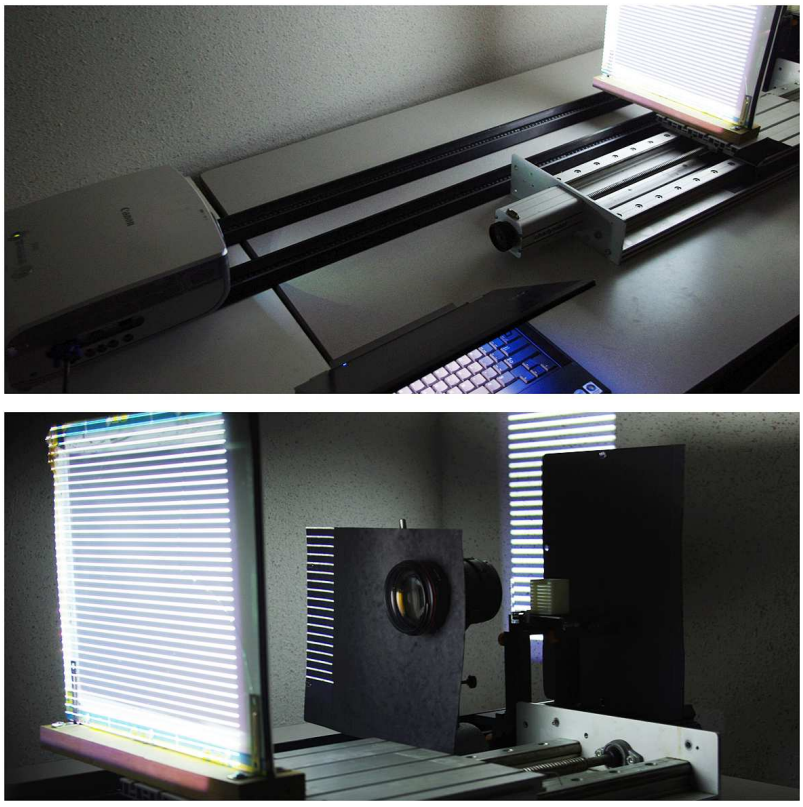


Figure 5.4: Experimental set-up of the first system (light pattern projection system)

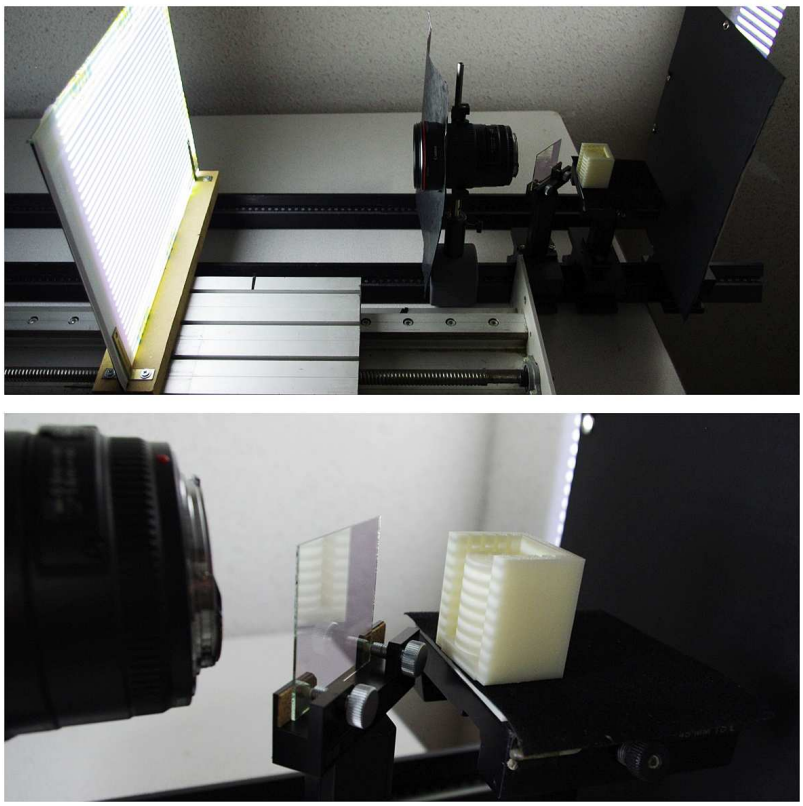


Figure 5.5: Experimental set-up of the second system (main optical system)

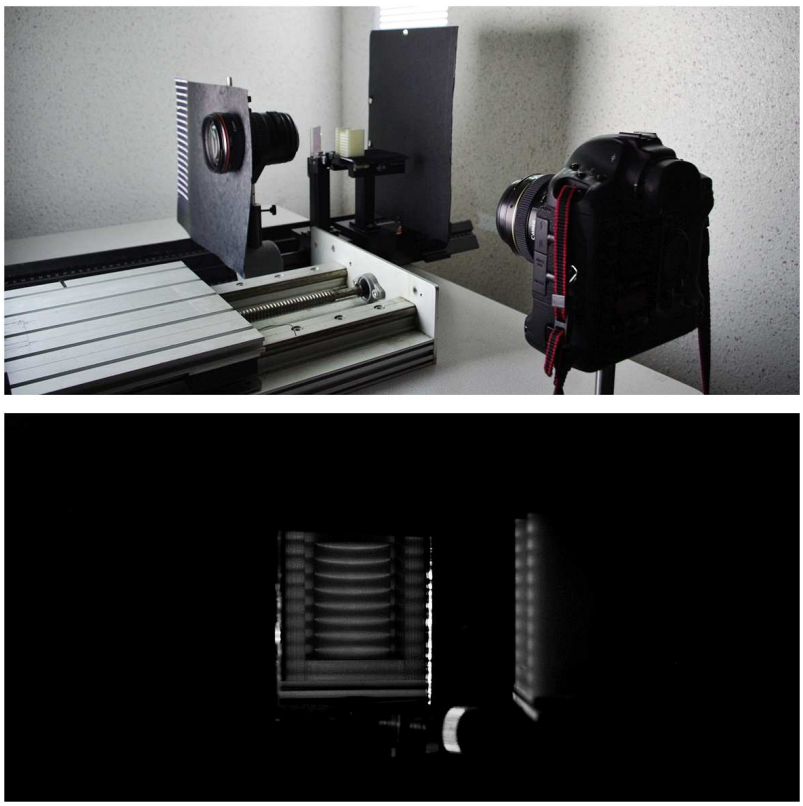
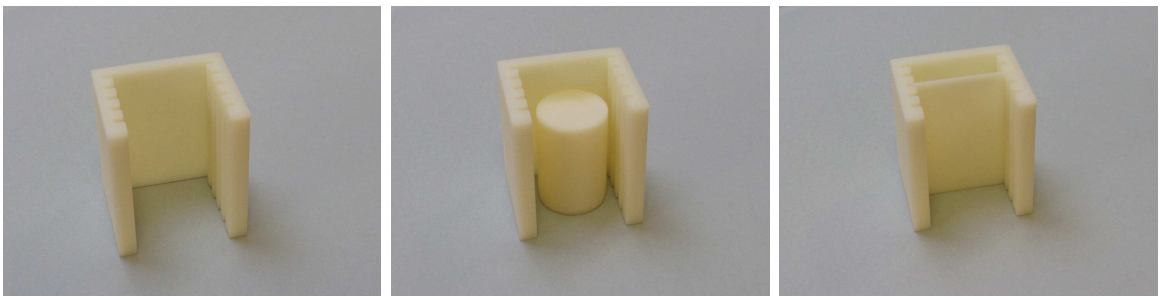


Figure 5.6: Experimental set-up of the third system (image acquisition system)

5.1.2 Calibration procedure

The object for the calibration is designed to have the slots to insert the planar planes corresponding to each depth as shown in figure 5.7 (a). It is also used as the fixed envelope for other test objects during the depth estimation process (see figure 5.7 (b)). We calibrate the system using a planar surface at 5 different depths ($D1 - D5$) as exemplified in figure 5.7 (c).



(a) (b) (c)

Figure 5.7: Object envelope

Each plane is captured through the projection of 6 screen displacements ($Pl1 - Pl6$), resulting in a total of 30 calibrated images. The captured images of the calibrated objects are illustrated in table A.1 in Appendix. Similarly, we apply the same procedure to the test object, and we obtain 6 images. Before starting the main algorithm, the acquired images are pre-processed to keep only the meaningful part of images. First, we convert image from the RGB to the grayscale level. Then, we crop the images, selecting only the effective areas, which are object-projecting regions within the beam splitter. Figure 5.8 exemplifies the image pre-processing of our test objects.

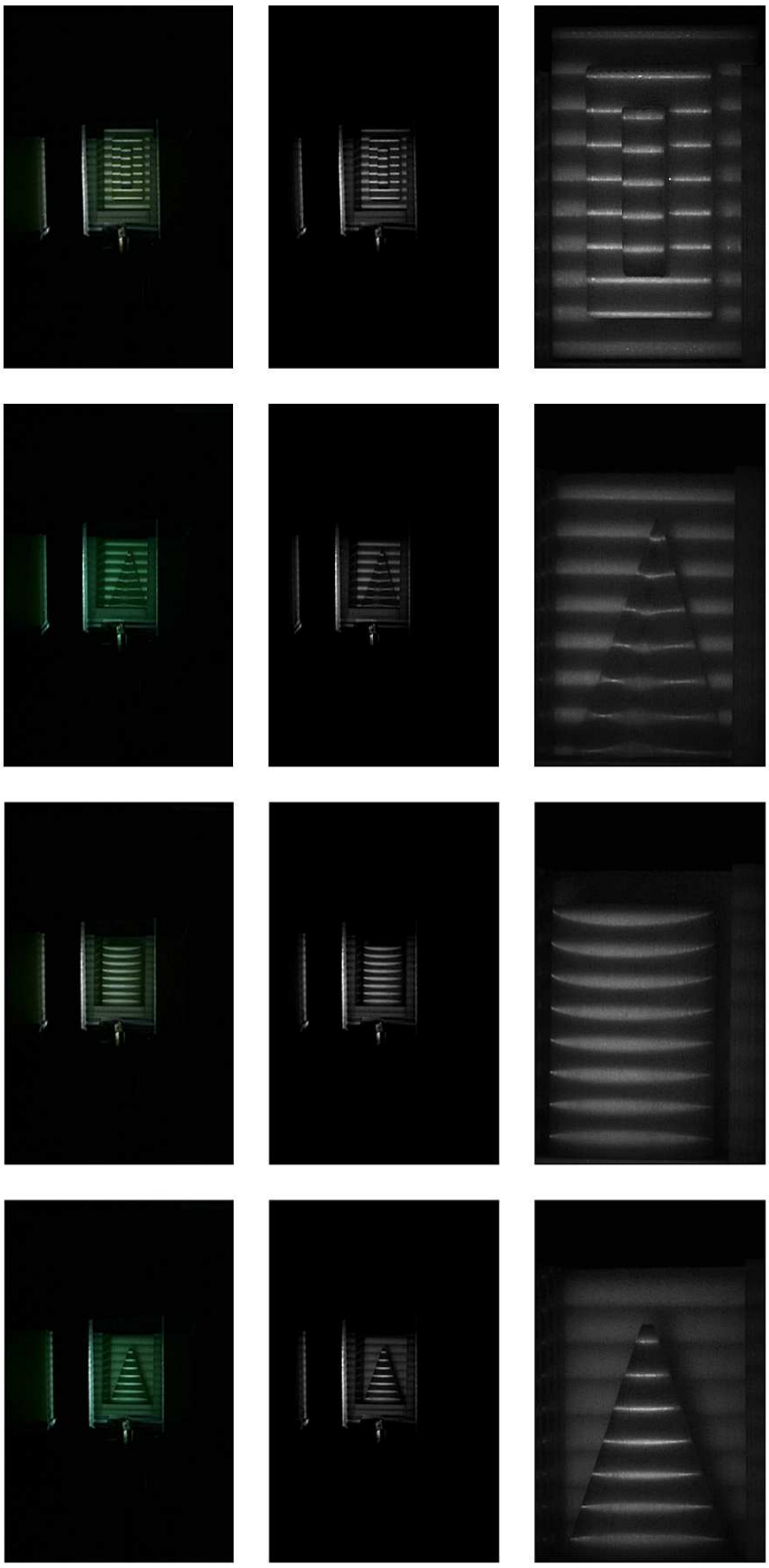
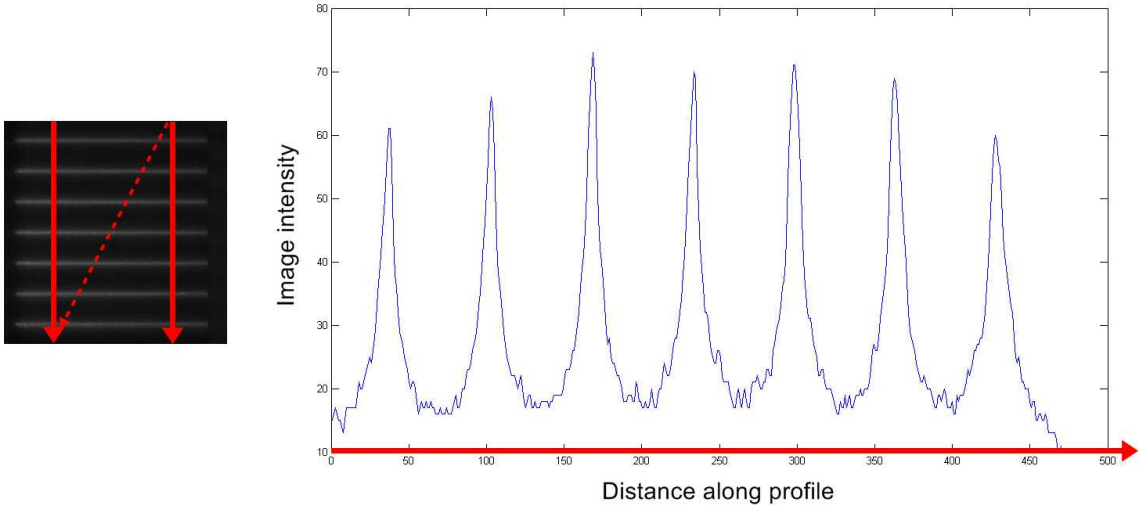


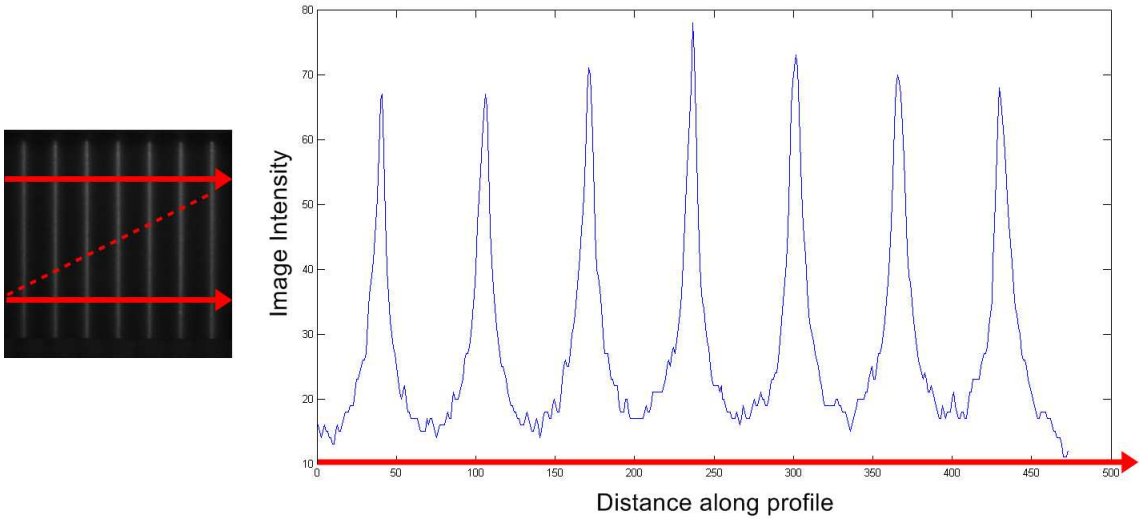
Figure 5.8: Image preparation from left to right: original image, grayscale image, cropped image

5.2 Image profile analysis

A 3D object can be thought of as the variation in depth over the object. However, these depth variations are missing during the process of imaging, and what remains is the intensity variations that are induced by the shape and the lighting. The profile analysis is performed, to extract the intensity values along multiline paths of the images. The algorithm computes equally spaced points along the specified path and uses interpolation to determine the image intensities for each point. This operation is performed along an orthogonal direction to the axis of the pattern projection as illustrated in figure 5.9.



(a)



(b)

Figure 5.9: Profile analysis in an orthogonal direction to the axis of the pattern projection (a) horizontal pattern (b) vertical pattern

To be precise, when the projected pattern is the horizontal stripes, the vertical profile analysis will be applied column by column, while for the vertical stripe pattern, the horizontal profile will be analyzed row by row. The output is stored in the profile stacks regarding their intensities and pixel coordinates. In either column-wise or row-wise approaches, profile analysis will provide numbers of peaks corresponding to the numbers of

stripe patterns. Each peak occurs at the center of its pattern and decays along both sides with a different speed. This scenario is based on the same concept that explains why the focused or sharp pattern gives a smaller width and higher profile intensity than the blurred pattern (see figure 5.10).

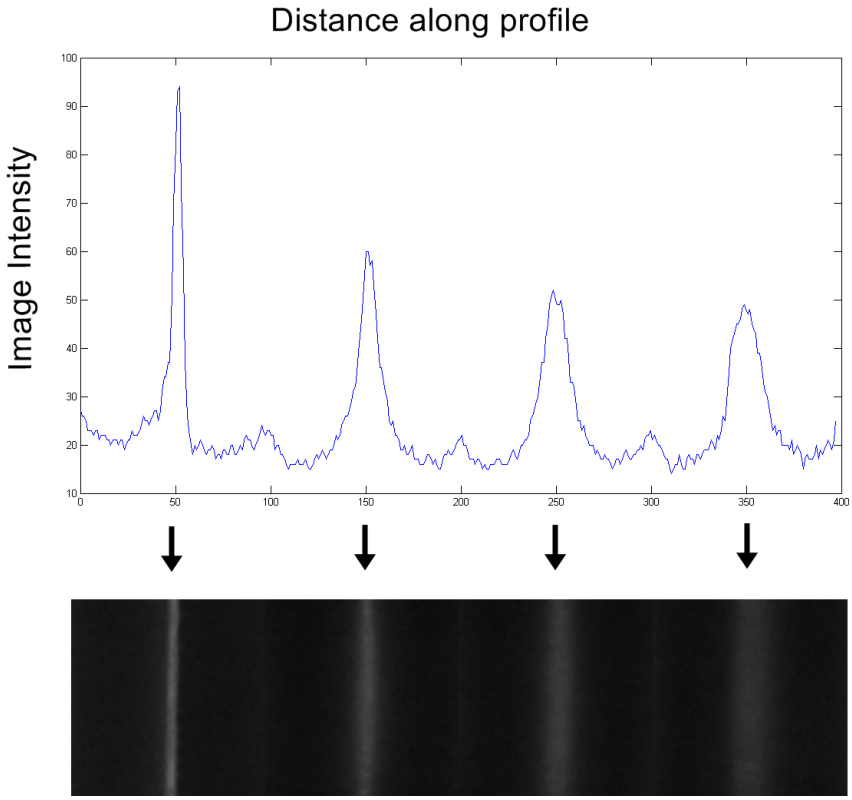


Figure 5.10: Different blur levels and their corresponding image profiles

5.3 Pattern localization

Each single profile contains either important data or noise. The difficult task is to differentiate the noise from the important signal before localizing the illumination pattern. The aim is to smooth the noisy part and to maintain the important intensity details simultaneously. Example of extracted intensity profile is shown in figure 5.11.

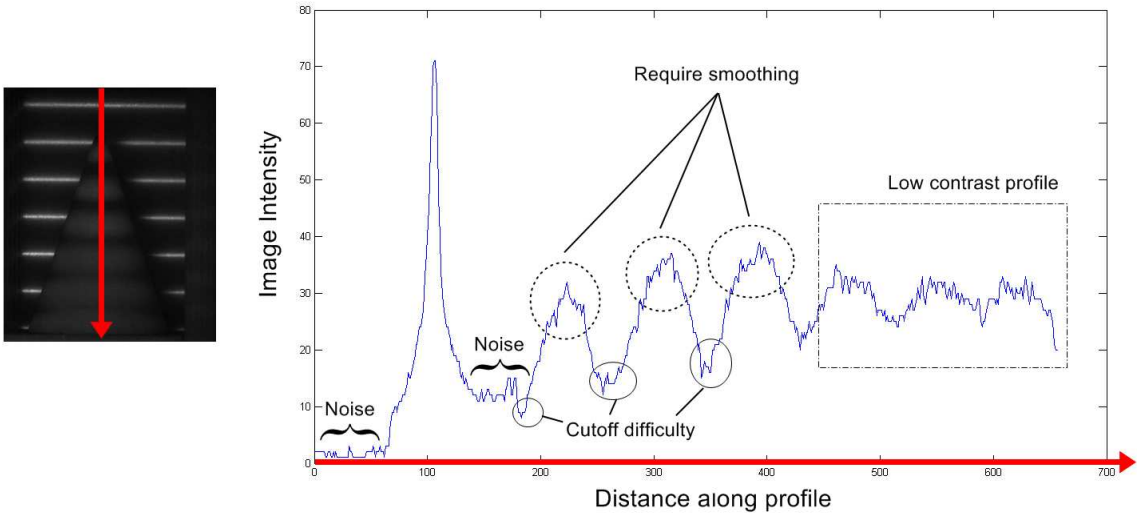


Figure 5.11: Example of extracted intensity profile

5.3.1 Profile smoothing

The smoothing of 1D signals is a well-known problem. We applied some of the most effective filters to the same sample of image profile. The comparison among different filters (Mean, Median, Wienner, and Sgolay filters) are shown in figure 5.12. In figure 5.13 (a), the result showed that the Wienner filter (in yellow) maintains the original intensity profile the most. While, the mean (in green) and the median (in magenta) filters smooth the profile more. However, our aim is to smooth only the cap of the profile to help the peak detection procedure. Nevertheless, it must maintain the width of each representing pattern profile because this is an important parameter used for our depth estimation. The width directly relates to the size of blur (spread parameter).

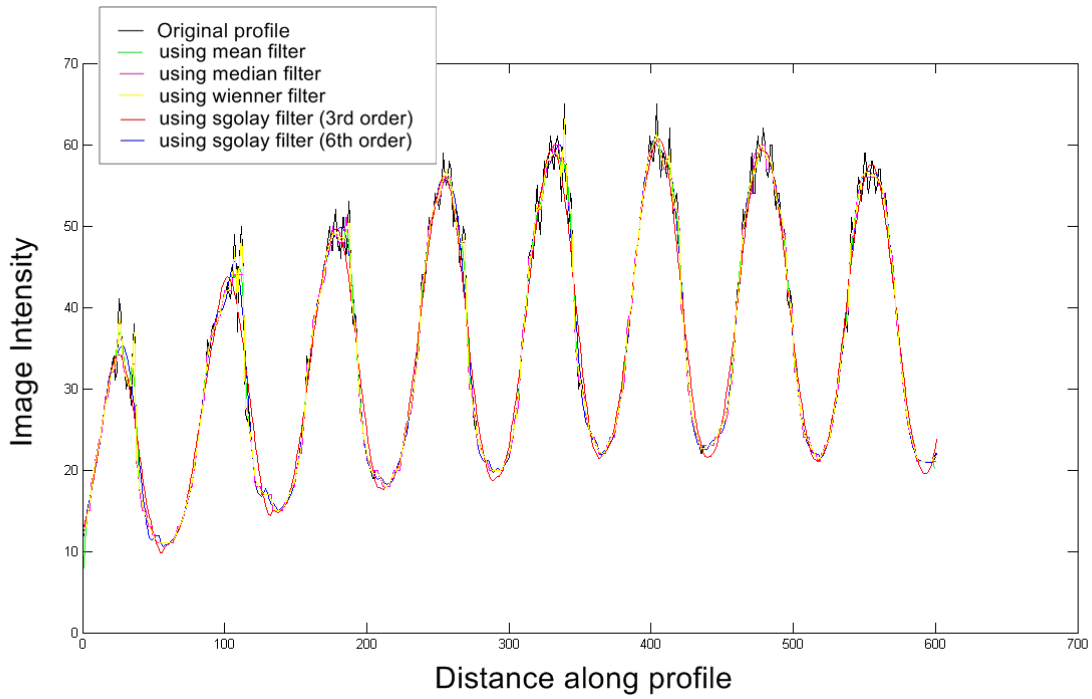
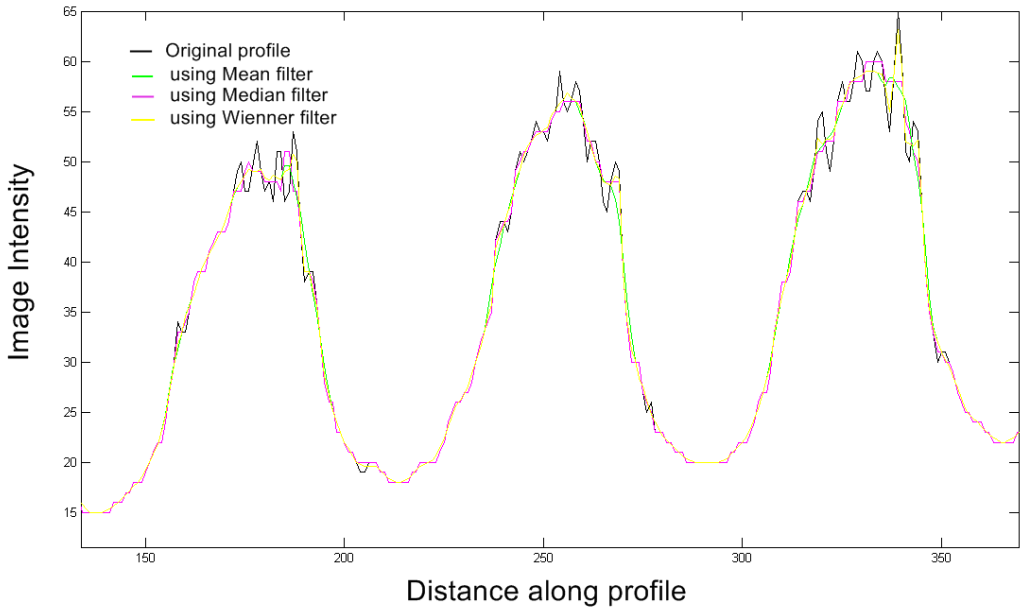


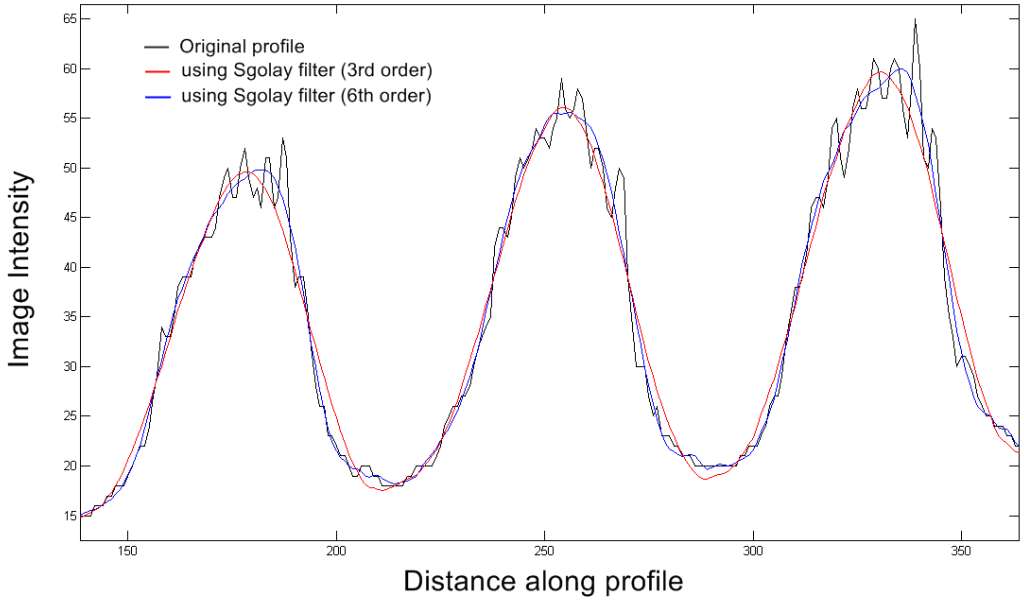
Figure 5.12: The comparison among different filters

Therefore, we selected the Savitzky-Golay(Sgolay) filter [92], which is a smoothing polynomial or a least-squares smoothing filter. It is very effective and suited for our type of signal, unlike typical Finite Impulse Response (FIR) filters that tend to filter out a significant portion of the signal's high frequency content along with the noise. The Sgolay filter essentially performs a local polynomial regression (of degree k) on a series of values (of at least k+1 points which are treated as being equally spaced in the series) to determine the smoothed value for each point. The Sgolay filtering algorithm allows us to define 2 parameters: frame size and polynomial order. The frame size indicates numbers of points to be evaluated each time. The polynomial order indicates how close the fitting to the original data. The higher degree polynomial makes it possible to achieve a high level of smoothing without attenuation of data features. In addition, the span must be odd, the polynomial degree must be less than the span, and the data points are not required to have uniform spacing. The results from 3^{rd} and 6^{th} orders are illustrated in red and blue lines, respectively. In our specific case, the 6^{th} order polynomial fits with 41 data points is preferable because it preserves features of the distribution such as relative maxima, minima and width. To be precise, we obtain perfect smooth at the cap, and also the closest

width to the original profile as desired. The mentioned effects can be clearly observed in figure 5.13 (b).



(a)

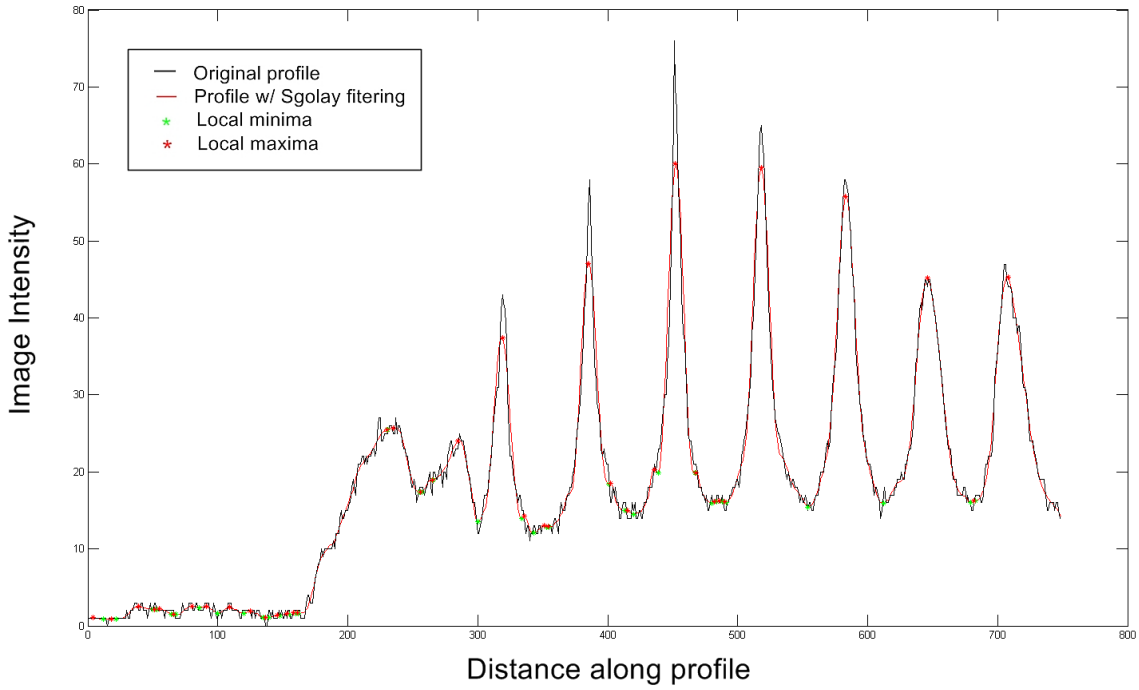


(b)

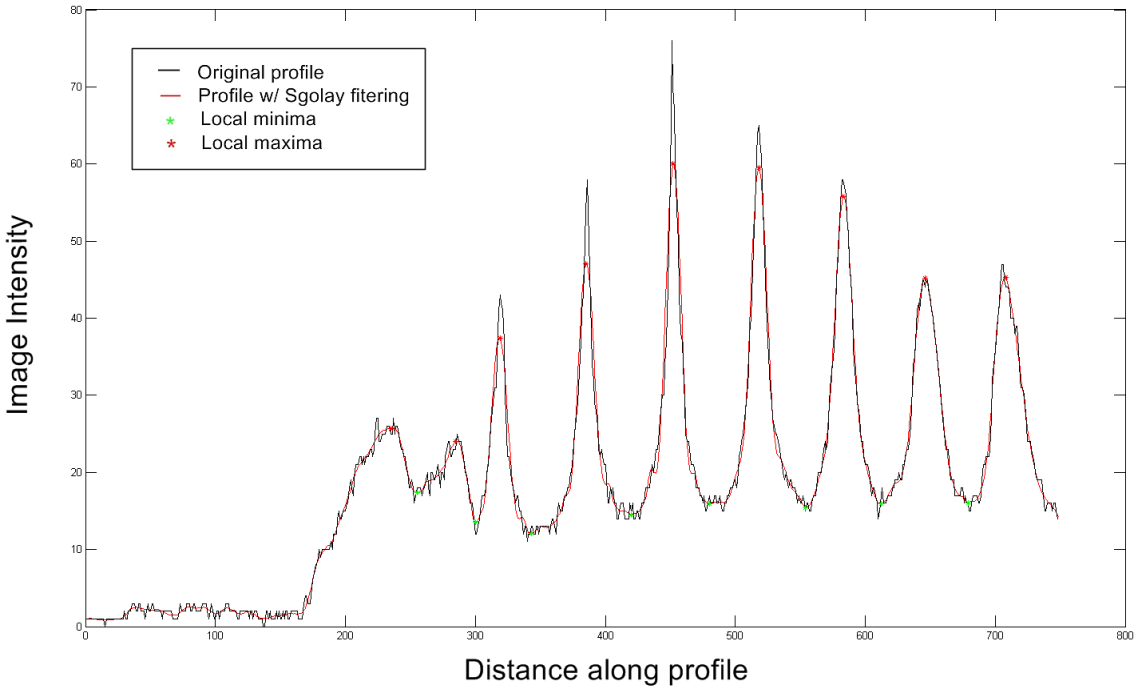
Figure 5.13: Effect of the smoothing filters

5.3.2 Cutoff determination

After smoothing the image intensity profile, the next task is to define a pattern cutoff coordinate. Therefore, the individual part of the profile can be fit with Gaussian model. We extracted the local maximum and minimum by an absolute peak detection algorithm. We prefer a non-derivative method because finding the zero-crossing of first derivative can yield false results in the presence of noise. This algorithm is adapted from the peakdet algorithm proposed by [93]. The strategy is to realize that a peak is the highest point between the valleys. To find one peak, it needs to have lower points around it on both sides. One constrain is the peak threshold, which is a minimum difference between a peak and its surrounding in order to declare it as the absolute peak. Similar scheme also goes with valley detection. The algorithm returns local maximum and local minimum with their respective positions. Figure 5.14 illustrates the comparison of the results with and without the absolute peaks determination.



(a)



(b)

Figure 5.14: Example of profile without (a) and with (b) absolute peaks and valleys determination

Moreover, instead of assuming the cutoff profile at the fixed ratio (e.g. 50% or 75% of the maximum), we compute another controlled parameter (α_1 and α_2) regarding the slope ratio between the local maxima and local minima on both spans. This approach is applied to guarantee that we optimize the cutoff portion of the significant information as much as possible (see also figure 5.15). The results can be separated into 3 cases as illustrated in figure 5.16, where the green and red marks indicate the local minima and local maxima, respectively. The blue and magenta marks indicate the controlled parameters at the distance away from the minima. The black dashed lines indicate the cutoff, where they are previously shifted to the higher intensity values. In addition, some profiles of the same depth can have different maximum intensities due to illumination variant over the object surface. Example of this position dependence problem is illustrated in figure 5.17. Therefore, the further computations have to always estimate the depth with respect to the individual location of the calibrated object.

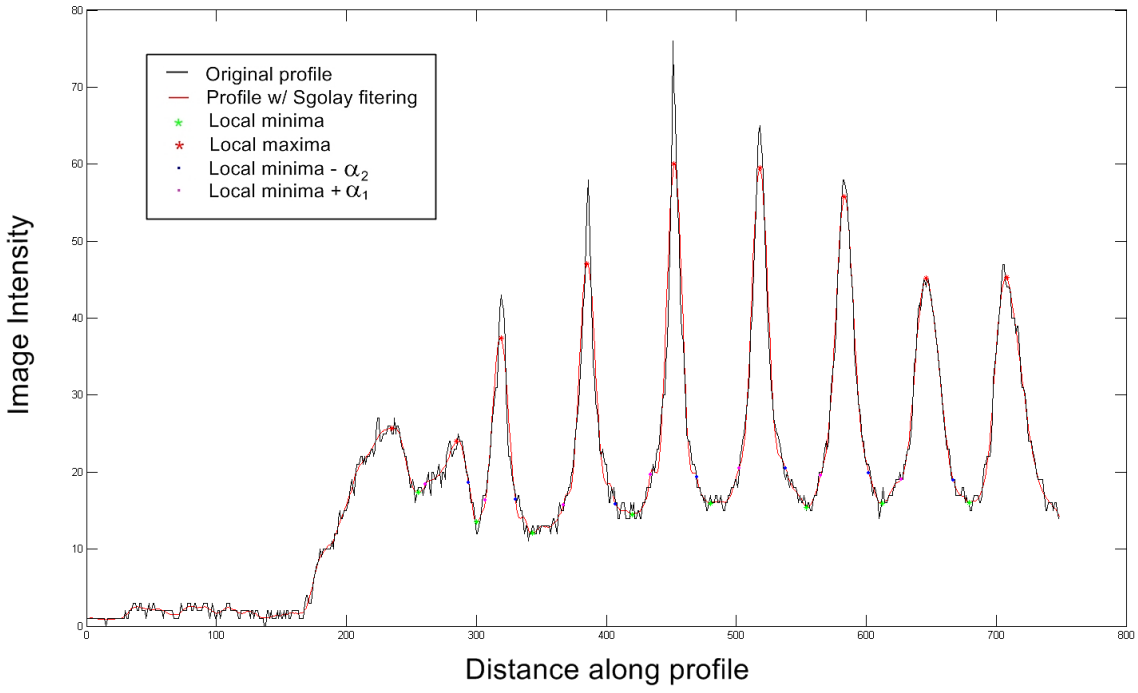
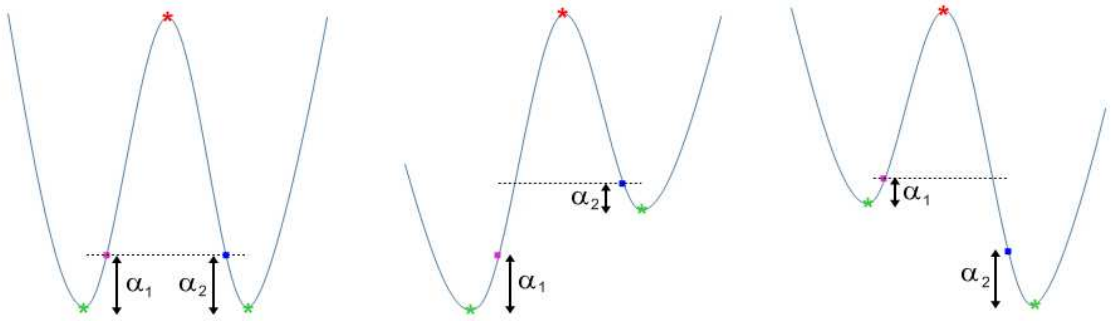


Figure 5.15: Example of profile with the controlled parameter



(a) (b) (c)

Figure 5.16: Possible cases of cutoff controlled parameter

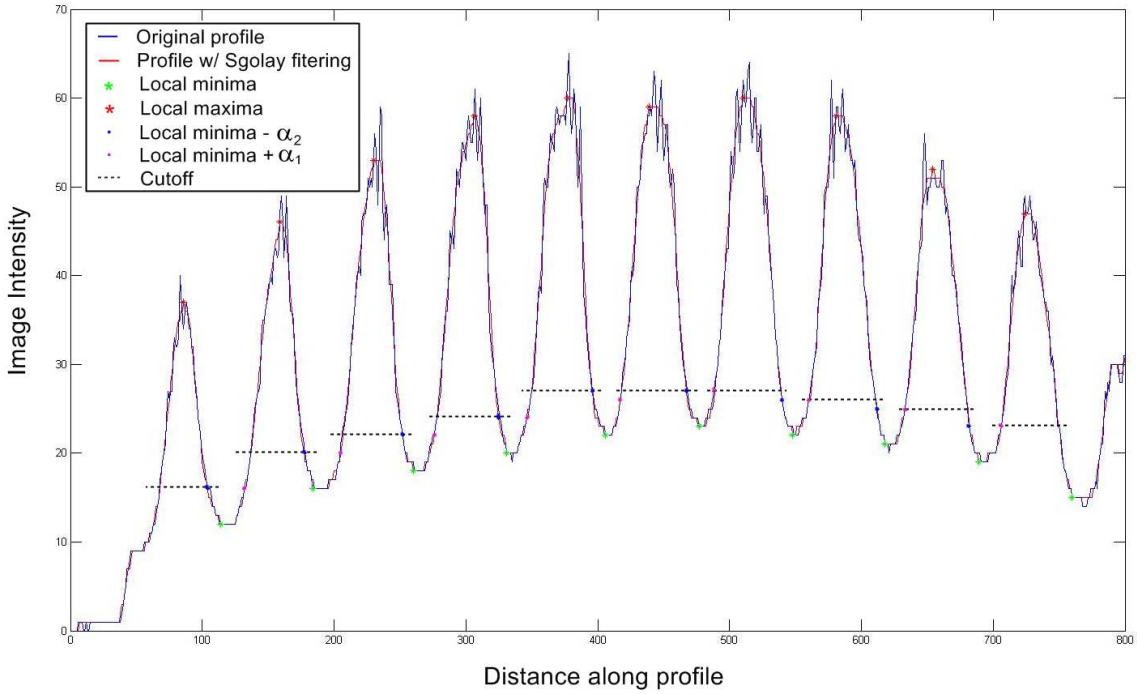


Figure 5.17: Example of profile with proper cutoff

5.3.3 Contrast Criterion

When the object sharpness varies over a wide range of depths, the blur levels in the far focus region are also considered as noise. To eliminate these high defocused profiles, a contrast criterion is also applied. The contrast threshold can have many possible values because there is a significant different between the important profile we want to keep,

and the low contrast profile we want to remove. Therefore, by setting this significant contrast, we can discard the defocused profile, in which the ratio between the maximum and minimum is lower than the contrast value.

it concludes the pattern localization process. We now obtain proper cutoffs to isolate the light pattern in each intensity profile. It is worth to mention that the criterion to determine the cutoff (e.g. maxima and minima detection, contrast thresholding, etc.) are computed using the profile from Sgolay filtering. However, once we obtain the cutoff, it is applied to the original image profile for the spread parameter calculation.

5.4 Spread parameter calculation

The distribution of light energy within the blur circle is referred to as the PSF. Because of the lens aberrations and diffraction effects, the PSF will be a circular blob, with its brightness falling off gradually rather than sharply. Thus, most algorithms use the two-dimensional Gaussian function instead of the Pillbox function. The fitting is done using the following function,

$$f(x) = A * e^{\frac{-(x-\mu)^2}{2\sigma^2}}. \tag{5.1}$$

Here, coefficient A is the amplitude, μ represents the average, and σ denotes the spread parameter. From the pattern patches isolated by the pattern localization, we determine their PSF individually. The spread parameters σ is extracted from the fitting between PSF and Gaussian model as exemplified in figure 5.18. The spread parameter is used to indicate the blur level in defocused images. Consequently, the depth can be deduced and assigned back to the pixel coordinates or the local maxima defined earlier. We iterate this algorithm for all of the light patterns that cover the whole object.

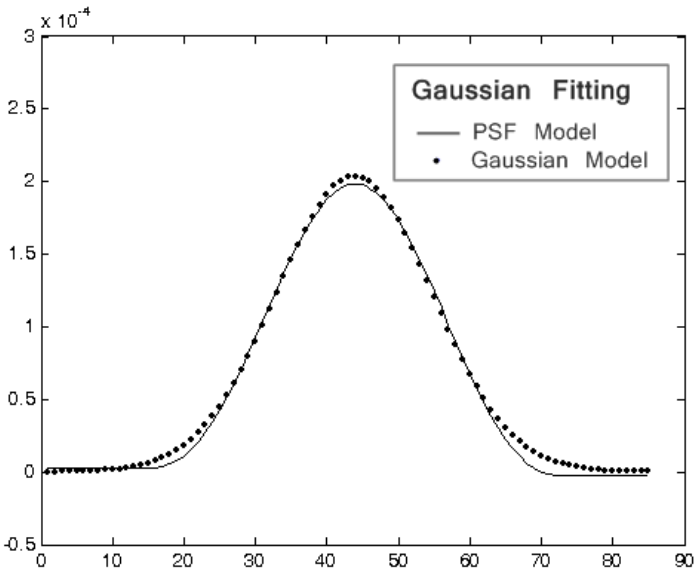


Figure 5.18: Example of Gaussian fitting

As mentioned earlier about the light pattern translation problem, the spread parameters σ of the light pattern obtained from each screen may shift, or may not locate at the same location. We solve the problem by defining a window, in which all the the spread parameters from the same light pattern are gathered. The size of window varies and is determined at every certain location of the image with respect to the number of light pattern. Therefore, for each point, we actually assume a small window containing all six screen estimates ($\sigma_1 - \sigma_6$). We do not select only one screen estimate but we do use all the six screen estimates within that window. Each of the screen estimates is then used to compute the final depth (following all steps as illustrated in figure 5.1) which is then assigned back to its pixel within the considered window.

5.5 Depth calibration

Depth calibration is performed only once by using 5 planar surfaces ($D1 - D5$) at 6 different screen positions ($Pl1 - Pl6$). Theoretically, there is no maximum numbers of screen position. Nevertheless, if many screen positions are used, less or no defocused

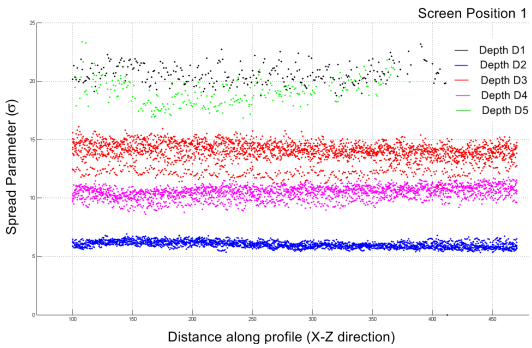
information is extracted from the blurs, and our approach will become the DFF. On the other hand, at least 3 screen positions are required to be able to determine the Parabola parameters described in section 5.5.3. Therefore, six screen positions are the compromised setting that ensures near-focus images are achieved for all the depths. We define the spread parameter at plane position i and depth j as $\sigma_{Pl_iD_j}$. The aim of depth calibration is to find the relationship among the available information (the depth, the extracted spread parameters, and the screen positions) for depth estimation. In the real experiment, the test object is captured from 6 screen displacements. Some preliminary approaches were investigated and summarized as the following.

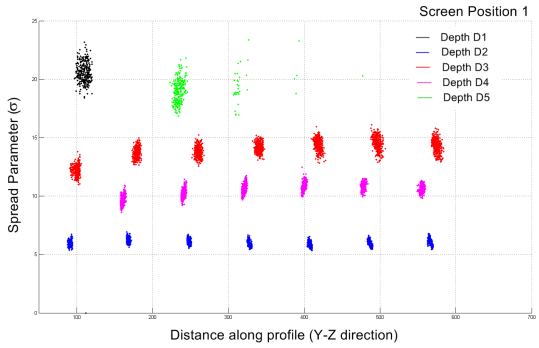
5.5.1 Solution 1: Relationship between spread parameter and plane position

Firstly, we analyzed the spread parameter of the calibrated depths at each screen position ($\sigma_{Pl_iD_{1-5}}$). From one screen position, we obtain different spread parameter values for each depth. Furthermore, for the same depth at different screen positions, the spread parameter values also vary. Then, we draw the depth plot (spread parameter versus the plane position) as shown in figure 5.3. We expected to obtain spread parameter of the test object somewhere linearly between the spread parameter of the calibrated depths. However, with the test object, the curve does not follow the calibrated depth plot. To be specific, there are no relationship and no repetition among the calibrated depths and therefore we cannot predict the final depth. For instance, let consider the screen position $Pl1$ in table 5.3. Depth $D4$ in (magenta) lie in between depth $D2$ in blue and depth $D3$ in red. This shows that the depths are not ordered. Meanwhile, for depth $D2$ (in blue), $D3$ (in red) and $D5$ (in green) from the screen position $Pl4$, the depths are ordered but still not linearly.

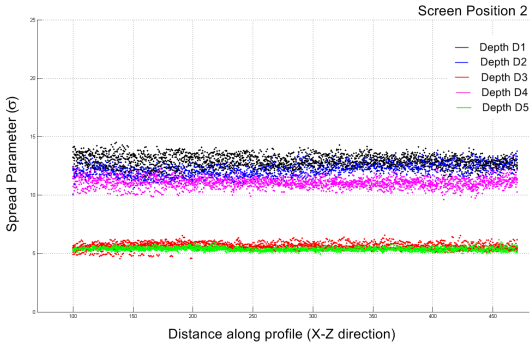
Screen Position	Spread parameter in X-direction	Spread parameter in Y-direction

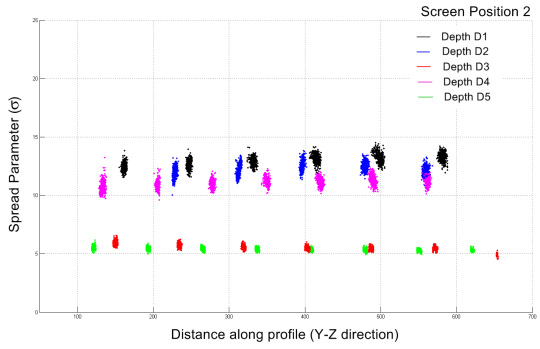
Pl 1



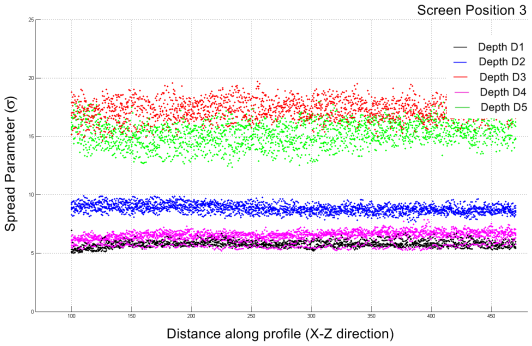


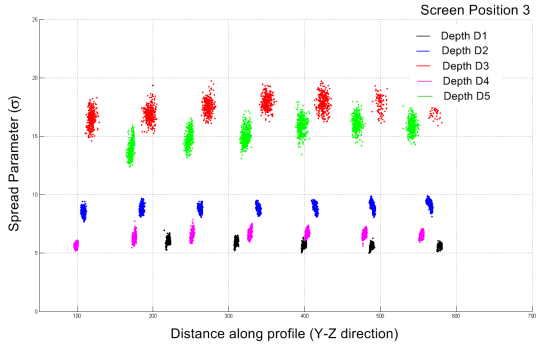
Pl 2



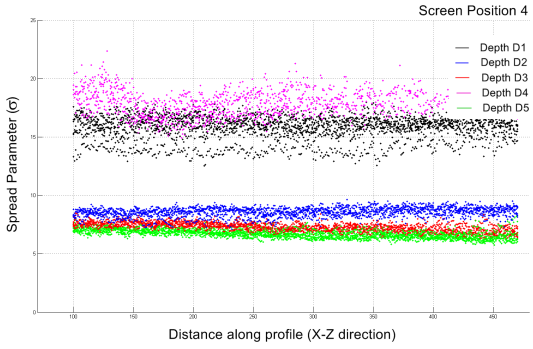


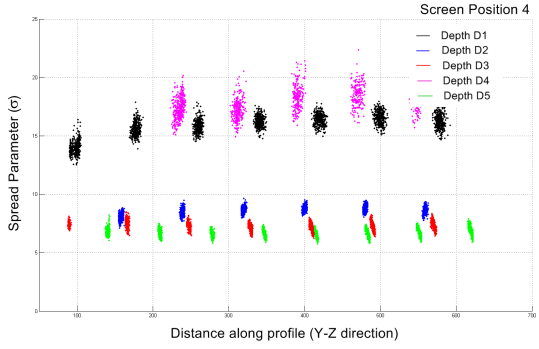
Pl 3





Pl 4





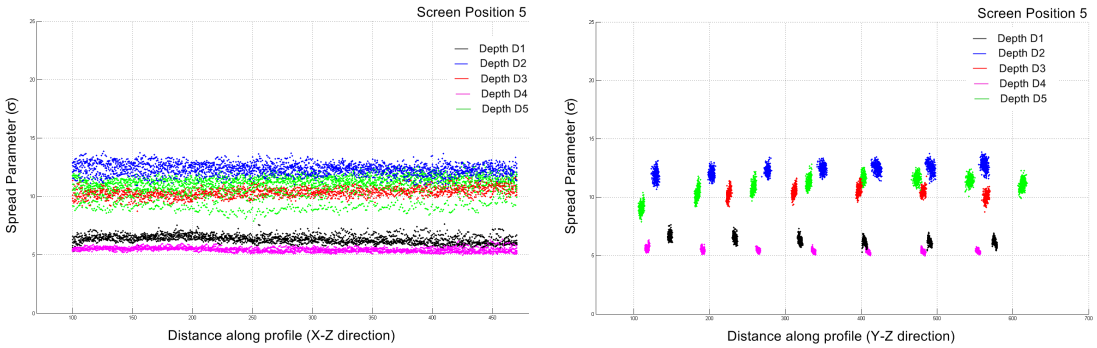
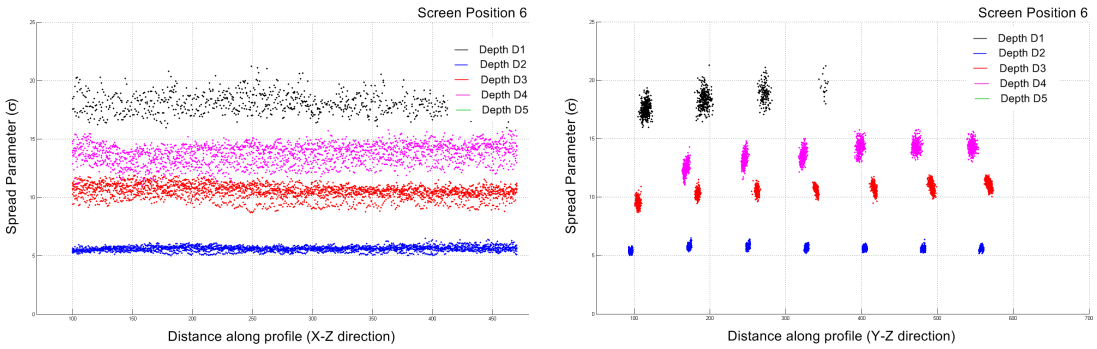
Pl 5	
Pl 6	

Table 5.3: Relationship between spread parameter and plane position

5.5.2 Solution 2: Confident weight map

This solution also uses the plot in figure 5.3. The strategy is to create the confident weight map to determine the final depth. We first calculate the mean value and standard deviation of the spread parameter at each depth. We assume that this relationship follows the *dsigmf* function (see figure 5.19).

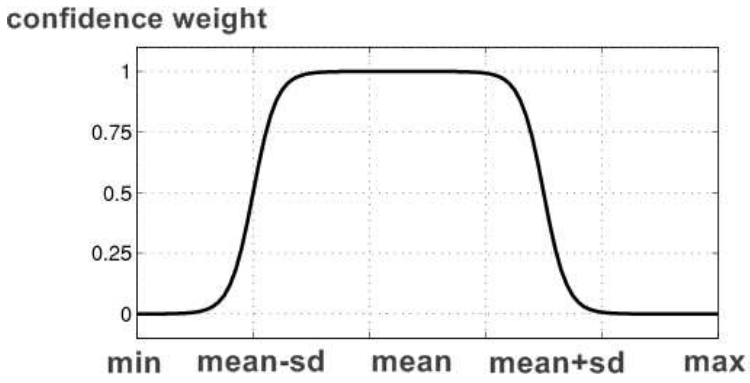


Figure 5.19: dsigmf function

The membership function *dsigmf* depends on four parameters, $m1$, $n1$, $m2$, and $n2$, and is the difference between two of these sigmoidal functions f_1 and f_2:

$$f(x;m,n) = f_1(x;m1,n1) - f_2(x;m2,n2) = \frac{1}{1+e^{-m1(x-n1)}} - \frac{1}{1+e^{-m2(x-n2)}}\ , \tag{5.2}$$

where

$$\begin{cases} m & \text{is the mean value of each depth } (m = m1 = m2), \\ n1 & \text{is the mean value subtracted by standard deviation, and} \\ n2 & \text{is the mean value added with standard deviation.} \end{cases}$$

However, our confident weight function W is slightly different,

$$W = \left(\frac{f(x;m,n)/f(m)}{(n2-n1)/(min(SD)\times 2)} \right)^t\ , \tag{5.3}$$

where

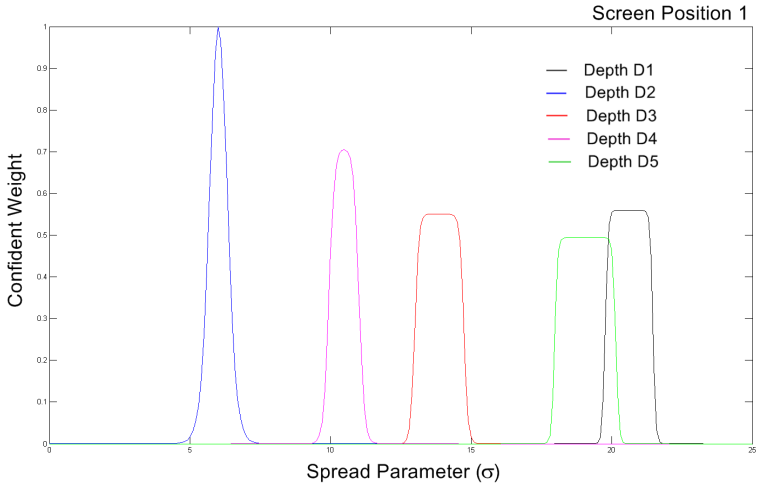
$$\begin{cases} SD & \text{is the standard deviation, and} \\ t & \text{is the tuned parameter for decay speed of the function (this case } t = 0.5). \end{cases}$$

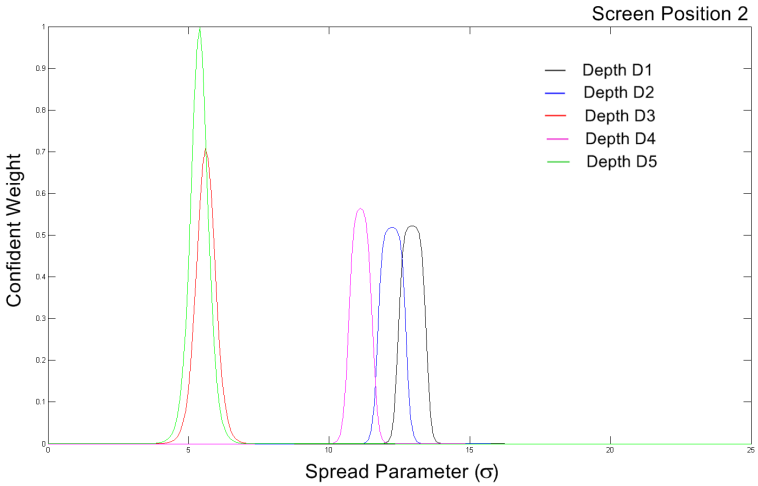
The term $\frac{n2-n1}{min(SD)\times 2}$ is used for the normalization according to the nature of blur. When the pattern is sharp ($\frac{n2-n1}{min(SD)\times 2} = 1$), it has high confident weight with small width. While,

for the blur pattern ($\frac{n2-n1}{min(SD)\times 2} > 1$), it has lower weight and larger width. The confident weight maps are illustrated in table 5.4, where the small spread parameters corresponding to the sharp pattern are given the higher confident weight and thiner width than the large spread parameters (blur pattern). We did not plot the confident weight map for screen position 6 since some spread parameter of calibrated depths cannot be extracted (too blurred). The depth is obtained by calculating the summation of all confident weights of each depth over all screen positions (W_j),

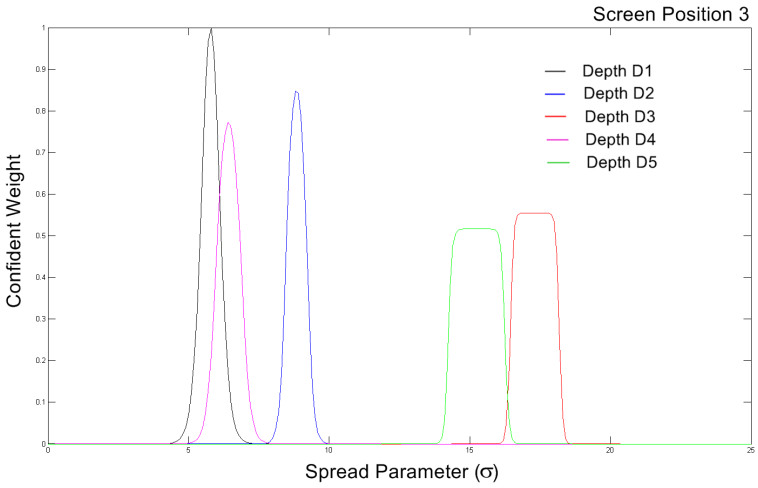
$$W_j = \sum_{i=1}^{6} W_{Pl_iD_j}. \tag{5.4}$$

The two highest weights W_j and their calibrated depth values D_j are then used to determine the final depth.

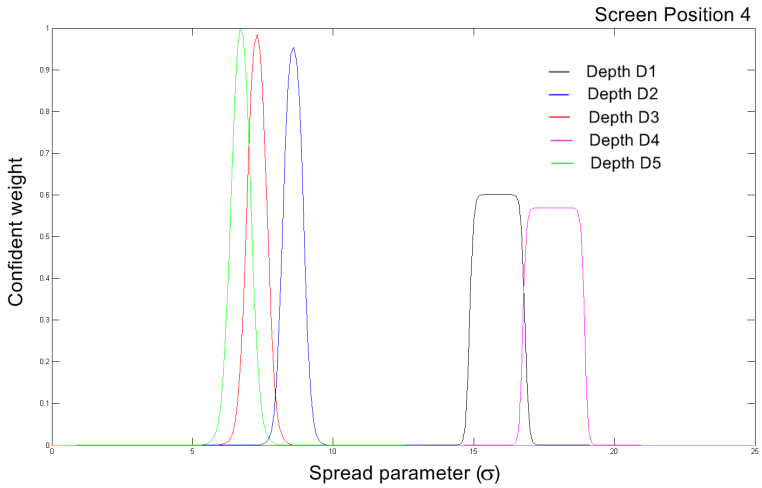
Screen Position	Confident weight map of calibrated depths
Pl 1	


Screen Position 2
Depth D1
Depth D2
Depth D3
Depth D4
Depth D5
Confident Weight
Spread Parameter (σ)

Pl 2


Screen Position 3
Depth D1
Depth D2
Depth D3
Depth D4
Depth D5
Confident Weight
Spread Parameter (σ)

Pl 3


Screen Position 4
Depth D1
Depth D2
Depth D3
Depth D4
Depth D5
Confident weight
Spread parameter (σ)

Pl 4

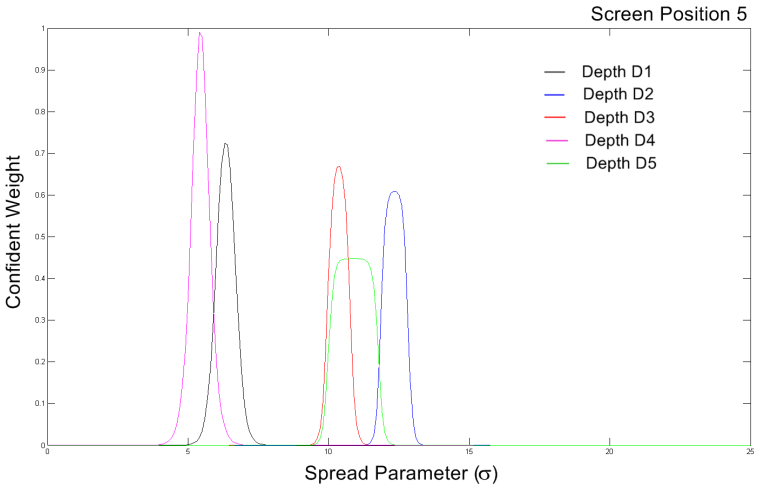
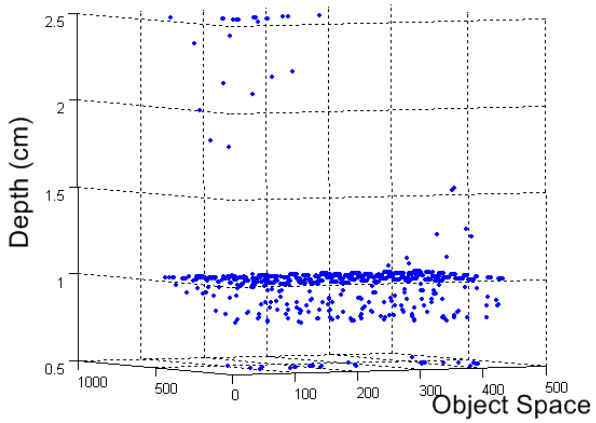
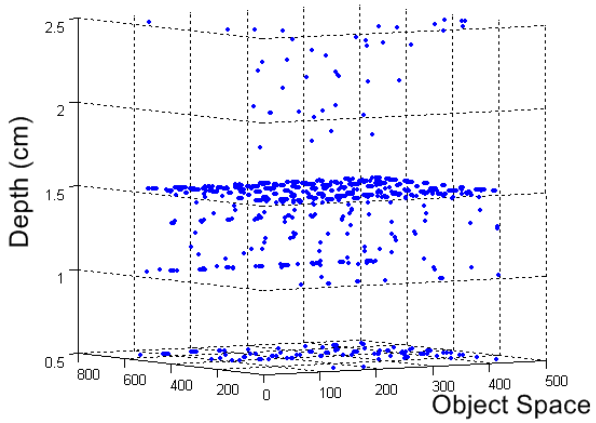
Pl 5	

Table 5.4: Confident weight map

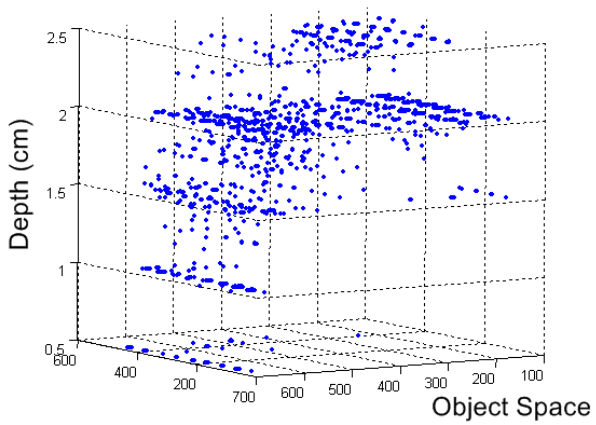
The point cloud obtained from this approach is illustrated in figure 5.20. The true depth for figure 5.20 (a) is 1 cm (b) is 1.5 cm, and (c) is 2.0 cm. We can see on the figures that numerous points are not correctly computed, which implies that the proposed approach still suffers inconsistency. The main problem came from the region where there are many depths overlapping. We cannot ensure that from which calibrated depth the weight refers to. In addition, it is very difficult to get good estimation for the intermediate depths because of the lack of order relationship between plane position and spread parameter (as seen in previous subsection).



(a) expected depth = 1 cm.



(b) expected depth = 1.5 cm.



(c) expected depth = 2 cm.

Figure 5.20: Example of results obtained from confident weight approach.

5.5.3 Solution 3: Relationship between spread parameter and depth

Our final solution is achieved by building the model defining the depth according to the value of the spread parameter σ at specific screen positions (Pl_i). In an ideal case, using equation 4.11, we can plot the relationship of the depth x against the spread parameter σ, as illustrated in Figure 5.21.

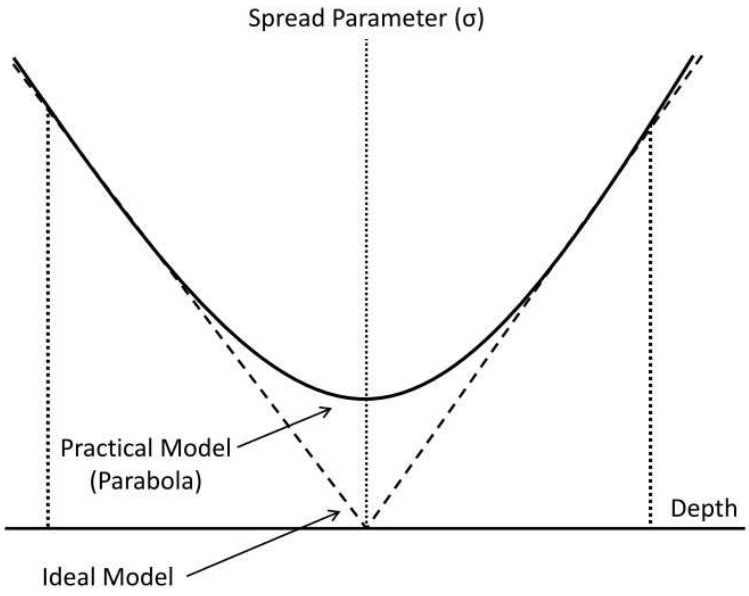


Figure 5.21: Ideal depth model and the practical conic model

This plot consists of 2 tangent lines having zero as a minimum at the center. Nonetheless, in practice, the elementary pattern has a minimum size on the captured image (see equation 4.6), causing a non-proportional law. To be specific, the spread parameter σ is not proportional to the object depth x for a small blurred interval around the center. Moreover, the optical system has certain acceptable sharp ranges related to the depth of field (DOF), resulting in a smooth valley instead of a sharp cut at the nadir. Therefore, we assume that the closed form model follows the Parabolic Function as follows:

$$\sigma_{Pl_i} = a_i x^2 + b_i x + c_i, \tag{5.5}$$

where a_i, b_i, c_i are the coefficient of parabolic function. Generally, if we have 3 unknowns and 3 equations, we can solve it as a linear problem. However, our system is dealing with more constraints. For each screen position, we calibrate 5 depths, which lead up to 5 equations. Therefore, with 3 unknowns (a, b, c) and 5 equations, we obtain an overdetermined system, which is solved in the least squares sense using the well-known Levenberg-Marquardt (LM) optimization algorithm [94]. The relationship between spread parameter and depth at a given screen position i is illustrated in figure 5.22.

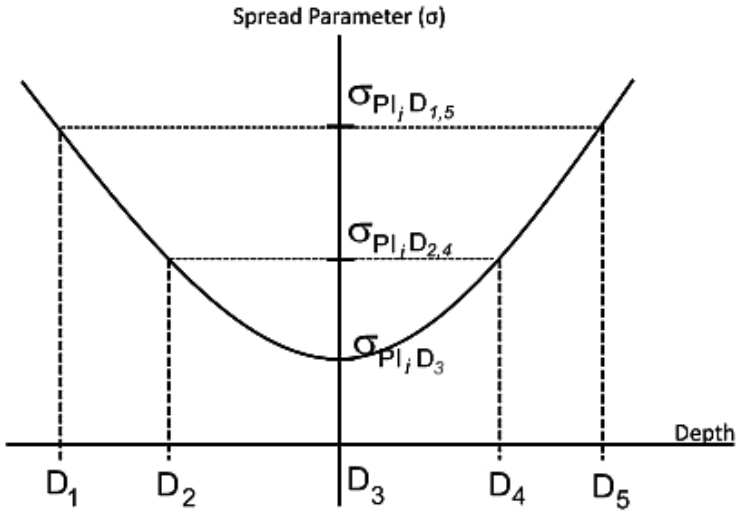


Figure 5.22: Relationship between spread parameter and depth

5.6 Candidate depth computation

For the real object used in the experiment, we follow the same manipulation for image acquisition, image profile analysis, pattern localization, and spread parameter calculation. Once we obtain σ, we can determine the candidate depths. By using the already extracted parabola parameters (a_i, b_i, c_i) from the depth calibration, we solve equation 5.5 for the final depth x. As seen in figure 5.23, it returns 2 possible solutions x_1 and x_2 for the depth, as follows:

$$x_1 = \frac{-b - \sqrt{b^2 - 4a(c - \sigma)}}{2a}, \tag{5.6}$$

$$x_2 = \frac{-b + \sqrt{b^2 - 4a(c - \sigma)}}{2a}. \tag{5.7}$$

Therefore, we need an additional clue to make the decision. The solutions to determine the final depth are explained in the following section.

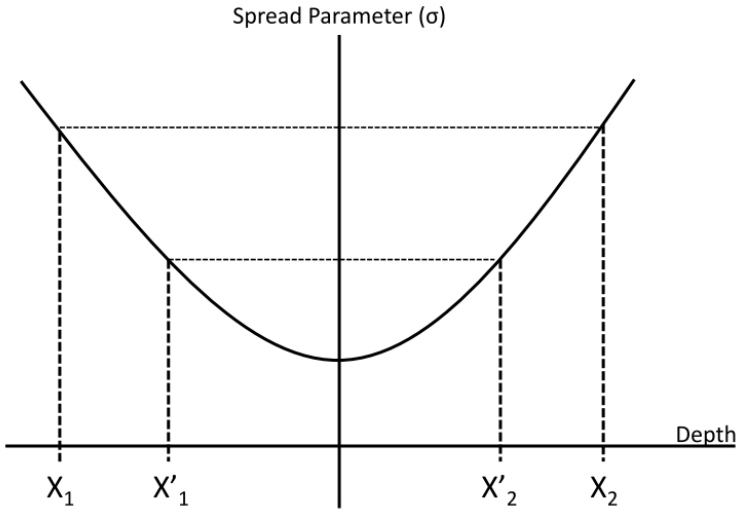


Figure 5.23: Candidate depths

5.7 Final depth evaluation

One spread parameter σ corresponds to 2 candidate depths. We present the successive attempts to determine the final depth.

5.7.1 Solution 1: Variance criterion

By assuming that the depths change gradually over small areas, we use the variance criterion to help in selecting the final depth. The variance is one of several descriptors of a probability distribution, which is used to measure the amount of the variation of variables by taking into account their average values and their probabilities. We first calculate 2 sets of candidate depths:

1. Matrix *LD* contains all the the lower candidate depth values x_1 computed from equation 5.6

2. Matrix *HD* contains all the the higher candidate depth values x_2 computed from equation 5.7

Then, we extract a window of size 3×3 from matrix *LD* and matrix *HD* as matrix L and matrix H, as shown in figure 5.24. We compute the variance of non-zero values within the matrix L and matrix H.

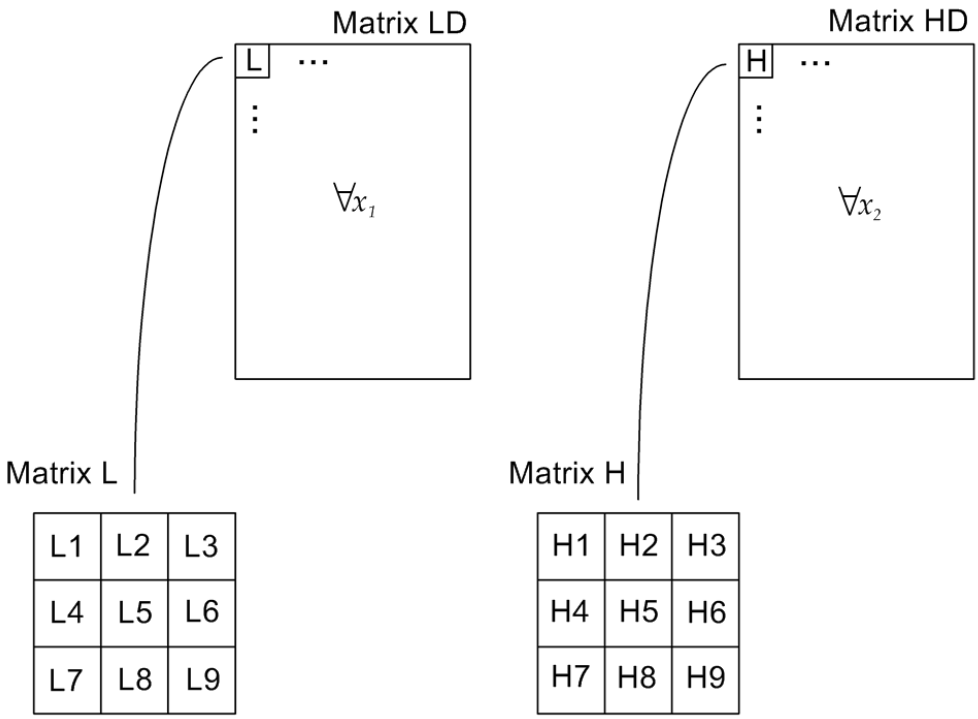


Figure 5.24: Illustration of the variance approach

$$var(L) = E[(L - \mu)^2] = \frac{\sum (L - \mu)^2}{N}, \tag{5.8}$$

$$var(H) = E[(H - \mu)^2] = \frac{\sum (H - \mu)^2}{N}, \tag{5.9}$$

where

$$\begin{cases} L & \text{is the non-zero variables within lower candidate depths window (i.e. } L1\text{-}L9) \\ H & \text{is the non-zero variables within higher candidate depths window (i.e. } H1\text{-}H9) \\ \mu & \text{is expected value (mean)} \\ N & \text{is numbers of non-zero variable} \end{cases}$$

Finally, the final depth x of the center pixel is selected from the candidate depth that belongs to the window of smaller variance.

$$x = \begin{cases} L5, \, var(L) < var(H) \\ H5, \, var(L) > var(H) \end{cases} \tag{5.10}$$

The algorithm iterates for all the pixels where candidate depths are located. Example of depth estimation by this approach is illustrated in figure 5.25. The candidate depths within both matrix are very close (small variance). In this case, the depth estimation is accurate. The correct depth of this position is 2.25 cm. Figure 5.26 shows the depth map of pyramidal object obtained from this approach. Numbers of correct final depths are achieved, but there are still lots of errors at the intermediate depths. Another problem comes from the position where both windows return high variance. Figure 5.27 exemplified this wrong selection case. The variance of the matrix H return smaller value than in the matrix L. Therefore, instead of the correct depth of 3.45 cm, the algorithm return 5.22 cm. Moreover, in the case where candidate depths exist in the window without neighborhood, this strategy is void (see figure 5.28). We now present the retained approach that works for all the cases.

Matrix L

2.2694	0	0
0	2.2572	0
0	0	2.2609

Var (L) = 3.913 x 10^{-5}

Matrix H

4.8839	0	0
0	4.8485	0
0	0	4.8496

Var (H) = 4.051 x 10^{-4}

Figure 5.25: Correct depth estimation from the variance approach

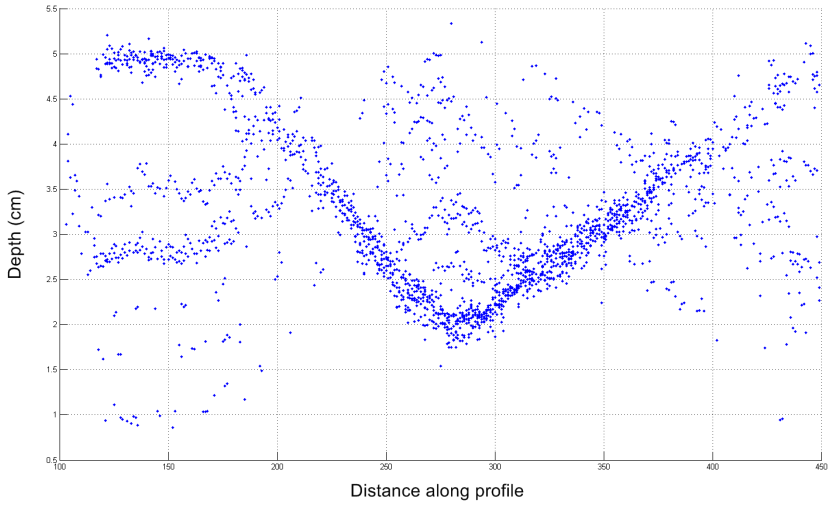


Figure 5.26: Depth map of pyramid object obtained from variance approach

Matrix L

2.7911	0	2.1224
3.5634	3.4518	3.5236
0	0	0

Var (L) = 0.3922

Matrix H

4.9550	0	4.9693
5.2719	5.2239	5.1876
0	0	0

Var (H) = 0.0221

Figure 5.27: Wrong depth estimation from the variance approach

Matrix L

0	0	0
0	L5	0
0	0	0

Matrix H

0	0	0
0	H5	0
0	0	0

Figure 5.28: Problem of candidate depths without neighborhood

5.7.2 Solution 2: Reference map

To select the correct depth value, we employ the reference map obtained from the spread parameter calculation during the calibration process. It is the mapping of the spread parameter at each screen position for all of the calibrated depths ($\sigma_{Pl_iD_j}$).

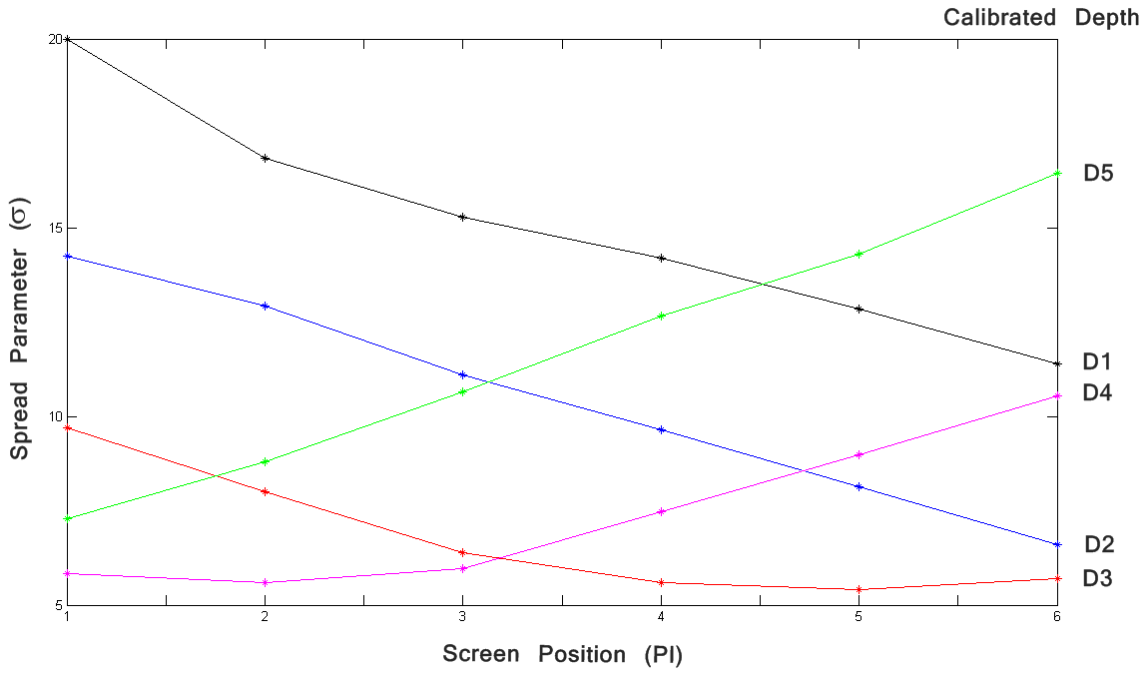


Figure 5.29: Reference map

By considering the reference map shown in figure 5.29, we plot the spread parameters of our object (to be reconstructed) according to their 6 observed screen positions (σ_{Pl_i}).

Then, we compute the minimum global distance, comparing them to 5 calibrated depths.

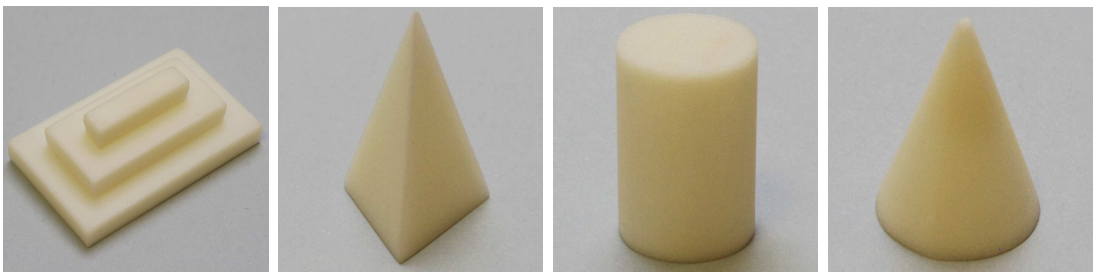
$$G_j = \sum_{i=1}^{6} \left\| \sigma_{Pl_iD_j} - \sigma_{Pl_i} \right\|, 1 \leq j \leq 5. \tag{5.11}$$

This step is to roughly define which calibrated depth our object belongs to. Eventually, the nearest candidate depth value (or $min(G_j)$) closest to the calibrated depth (D_j) will be selected as the final depth. With this approach, we obtained the final depths for all existing candidate depths and solved all the problem found earlier.

5.8 Experiment

5.8.1 Set up detail

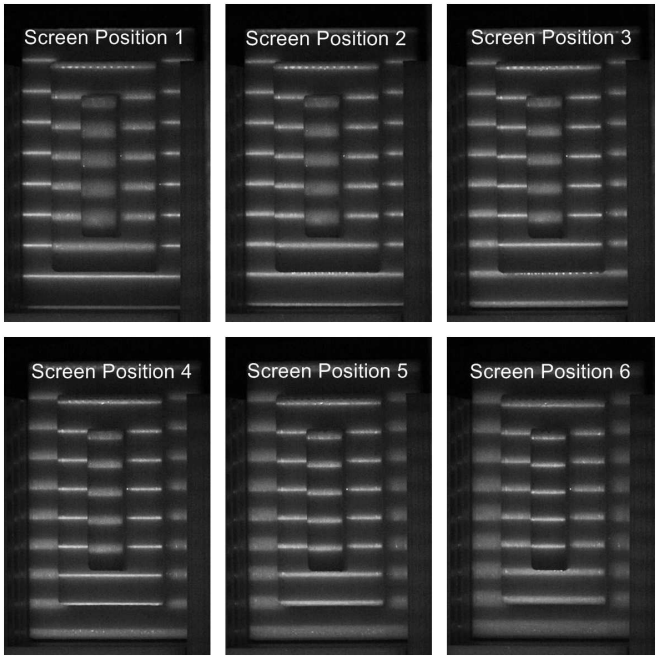
We conducted experiments using an acrylonitrile butadiene styrene (ABS) plastic built by a 3D printer as the test objects. The model of the 3D printer is a Dimension Elite produced by the Segway Inc. manufacturer. Without coating, the material has some reflectivity. Four types of surfaces have been tested: a planar, a pyramidal, a cylindrical, and a conic structure (see figure 5.30), they are all fitting in a bounding box of $2.5 \times 4 \times 6$ cm.



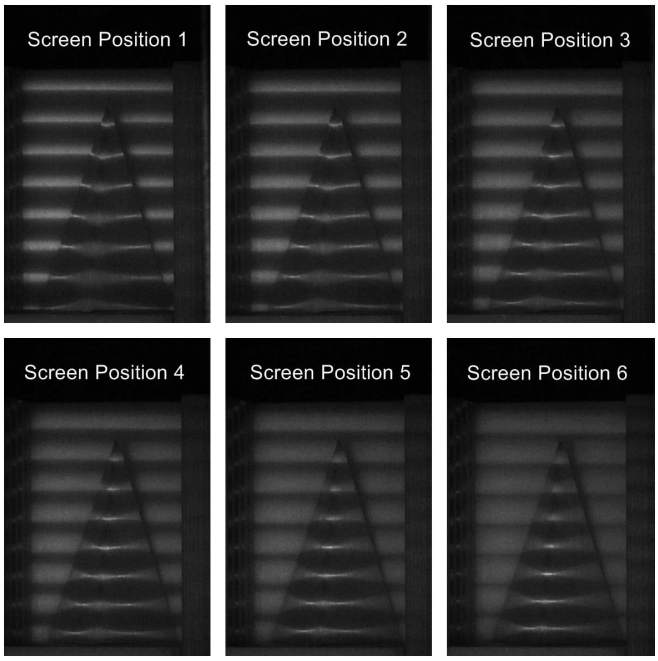
(a) (b) (c) (d)

Figure 5.30: Scene objects: (a) staircase, (b) pyramid, (c) cylinder, and (d) cone.

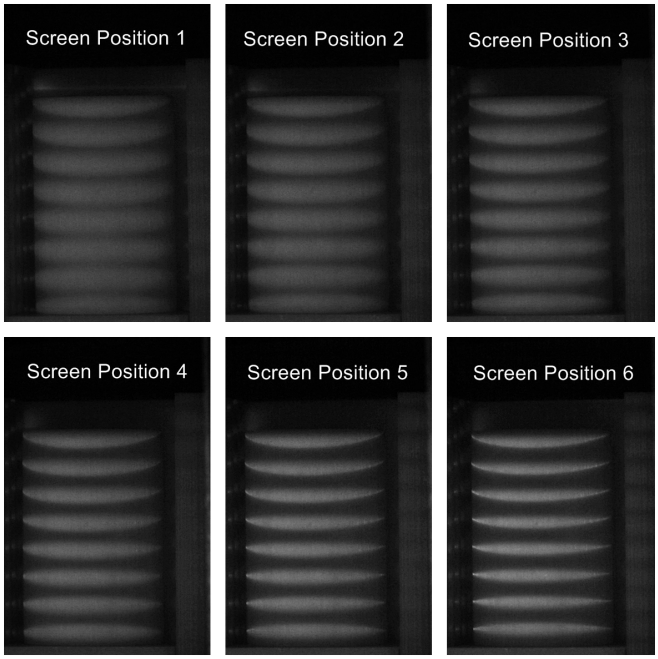
A sequence of light pattern images is shown in table 5.5. They are acquired at different projecting distances (physical displacement of 1 cm between each screen position). Because of some constraints of our optical setup (e.g. telephoto lens distortion and beam splitter size), the effective reconstruction areas are limited only at the center of the beam splitter.

Test objects	Captured Images from different screen Positions
Staircase	

Pyramid



Cylinder



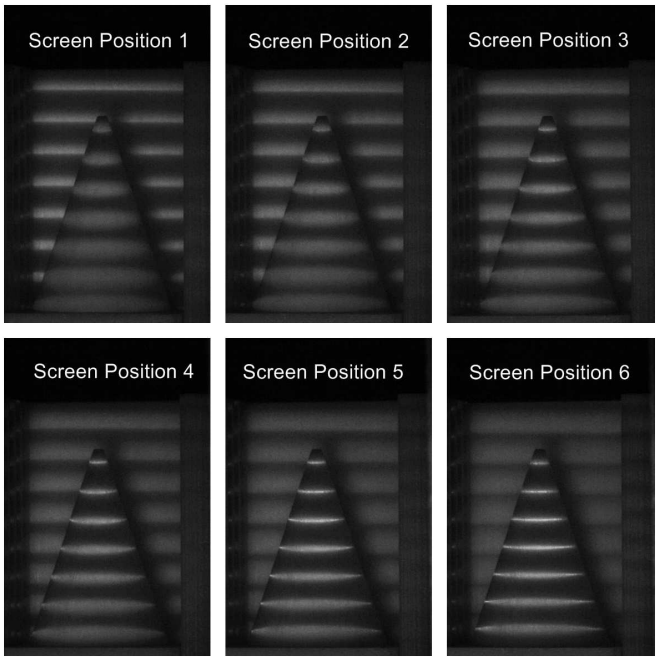
Cone	

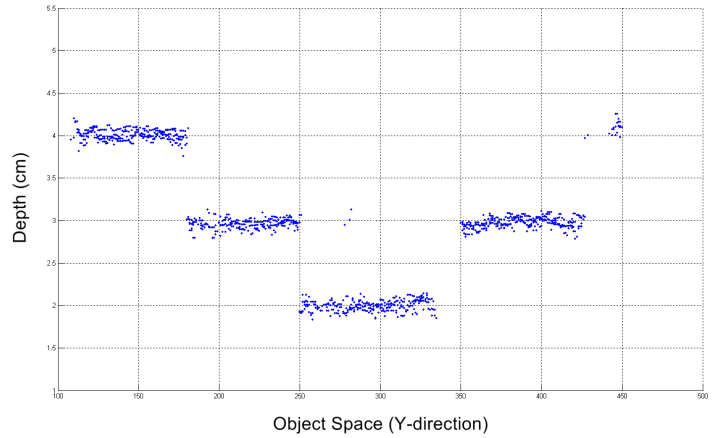
Table 5.5: Captured images of the test objects

The input images are then put into the stack for profile analysis and pattern localization. For each isolated light pattern, the spread parameter is extracted by fitting the PSF to the Gaussian model. The candidate depths are then calculated by equation 5.6 and equation 5.7 with the already known parabolic parameters from the calibration process. Eventually, the final depth is determined by taking the reference map into consideration.

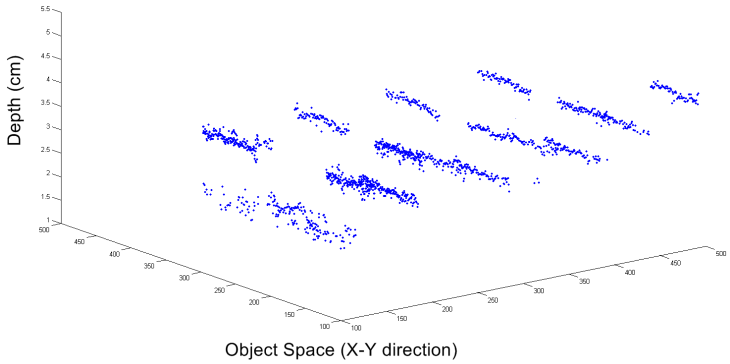
5.8.2 Results

The rough 3D model presents some preliminary results that are obtained from our implementation. The depth map illustrated in Figure 5.31 demonstrates the effective performance of the method in the case of the planar structure. A staircase object has a minimum depth at 2 cm, increasing on both sides by 1 cm until the maximum depth of 4 cm is reached. The result shows that both the real object depth and our estimated depth lie within close proximity. Figure 5.32- 5.34 illustrates the depth map of the pyramidal,

cylindrical, and conic objects, respectively, in which we can also retrieve the intermediate depths that do not exist during the calibration process.

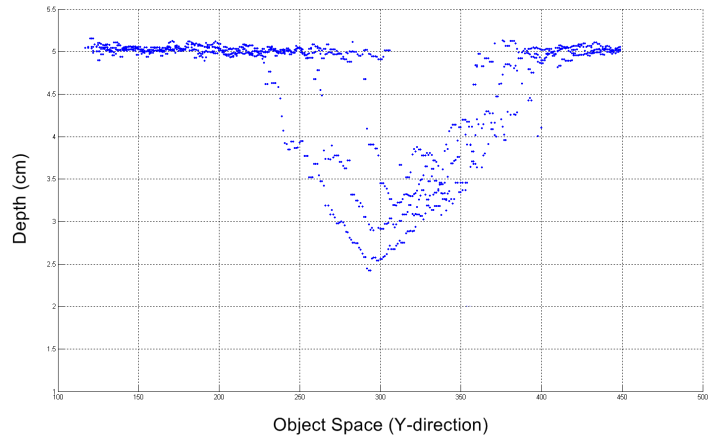


(a)

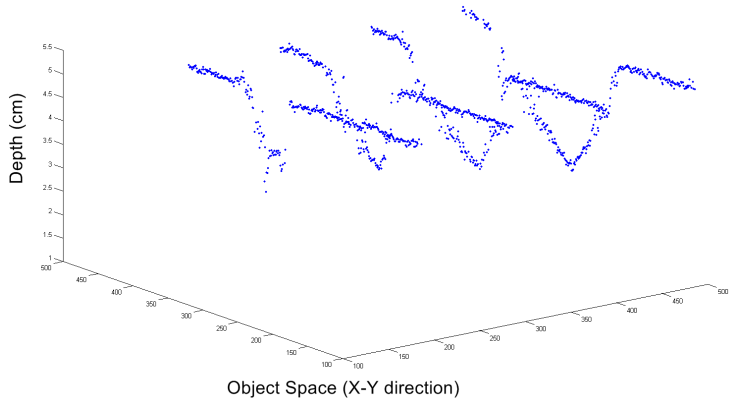


(b)

Figure 5.31: Depth map of staircase object

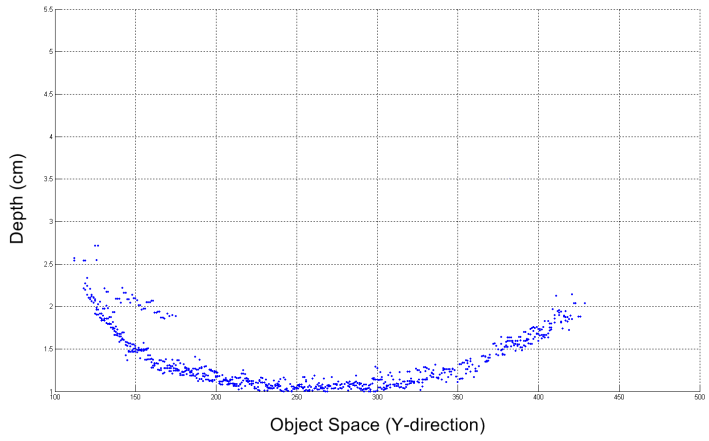


(a)

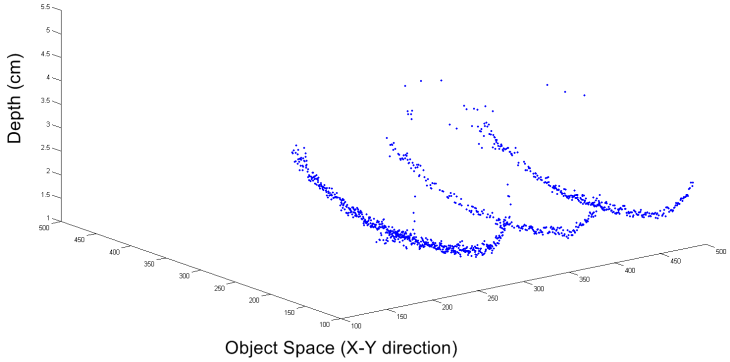


(b)

Figure 5.32: Depth map of pyramidal object

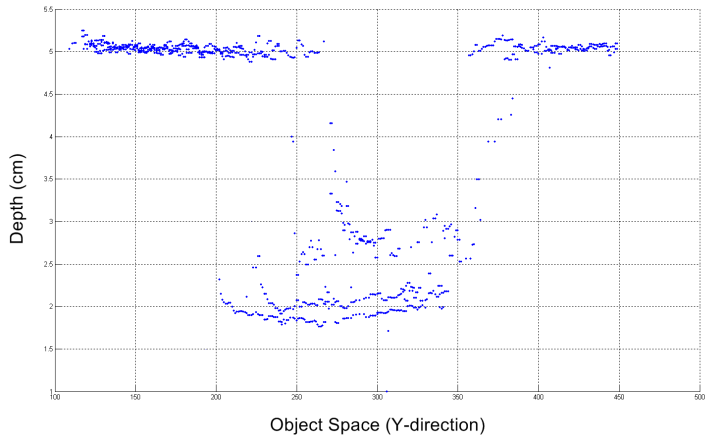


(a)

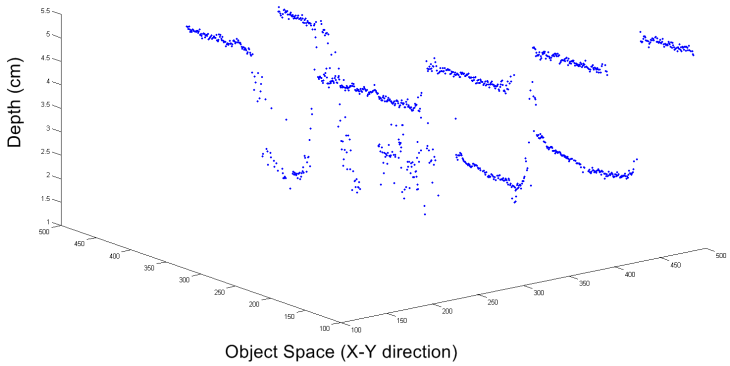


(b)

Figure 5.33: Depth map of cylindrical object



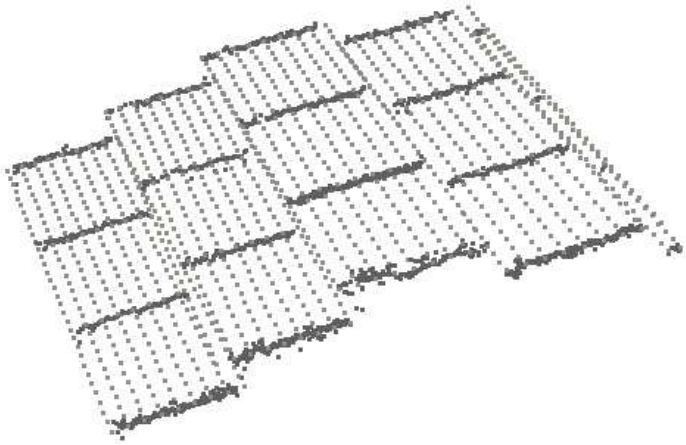
(a)



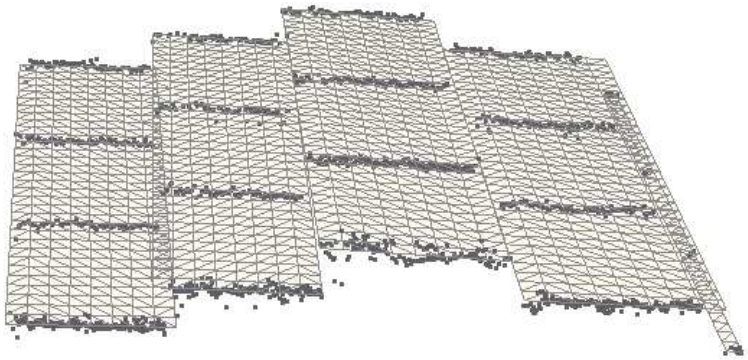
(b)

Figure 5.34: Depth map of conic object

By combining the point cloud acquired from the experiment and the pre-defined geometric shapes, we can perform the 3D modeling and surface fitting using the Rapidform software. The point cloud and the 3D reconstruction of the test objects are shown in Figure 5.35 - 5.38. The objects obtained by fitting are within the same dimensions than the real objects.



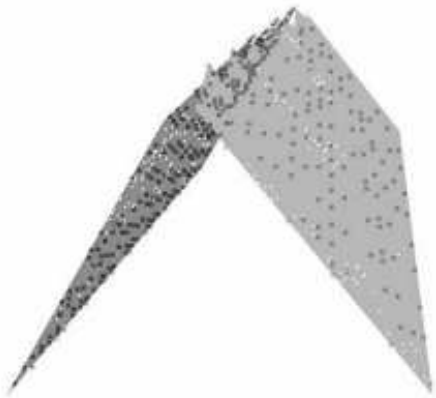
(a)



(b)

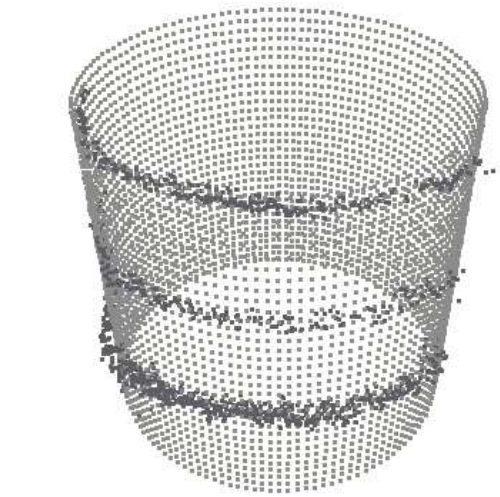
Figure 5.35: Rough 3D reconstruction of staircase: point cloud (a) and surface fitting (b)

(a)

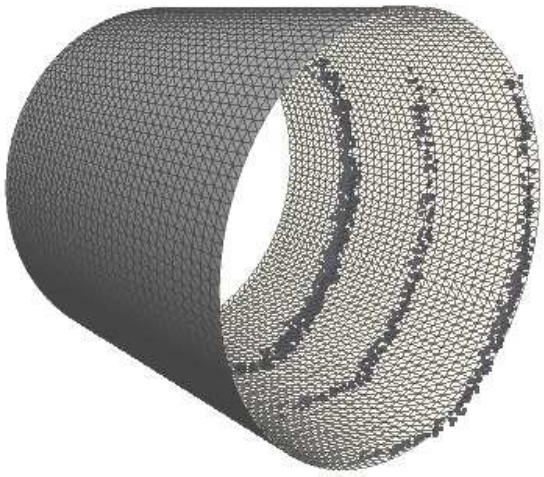


(b)

Figure 5.36: Rough 3D reconstruction of pyramid: point cloud (a) and surface fitting (b)

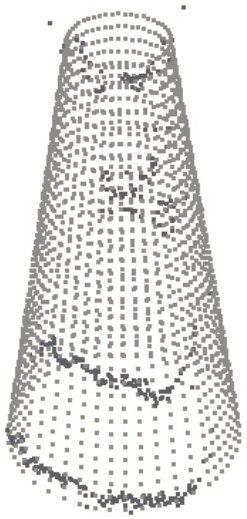


(a)

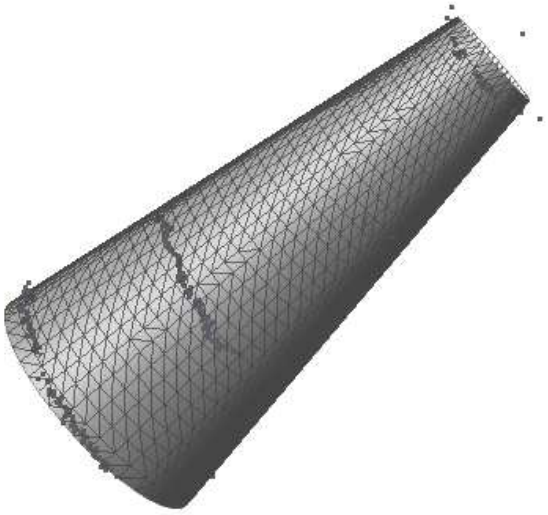


(b)

Figure 5.37: Rough 3D reconstruction of cylinder: point cloud (a) and surface fitting (b)



(a)



(b)

Figure 5.38: Rough 3D reconstruction of cone: point cloud (a) and surface fitting (b)

The results are compared to the actual depth provided for the 3D printer. The quantitative evaluations are presented in table 5.6. Both test objects have an average error of less than 0.35 mm.

	Staircase	Pyramid	Cone	Cylinder
Numbers of Point cloud	8784	6375	6320	4967
Average Error (mm)	0.17780	0.28429	0.33697	0.3464
Standard Deviation	0.28609	0.34347	0.2170	0.24719
Computational Time (sec)[a]	31.901	19.150	12.843	14.908
Computational Time (sec)[b]	208.099	203.790	182.135	201.360

Table 5.6: Experimental results.

[a]excluding offline profile analysis and calibration
[b]offline profile analysis and calibration

Denser point clouds and higher quality 3D reconstruction can be obtained once the variation and the number of projected light patterns increase. The total process can be time-consuming. With non-optimized Matlab code, the program takes less than one minute of computational time (on the machine equipped with a core 2 duo 2.2 GHz), excluding profile analysis and calibration. However, the benefit of our approach is that most of the processes that require a long time computation are offline processes. The calibration is required only once for a certain object material. Therefore, if we already have calibration data for a set of object materials, we will be able to create a 3D model quite quickly, as reported in table 5.6.

In addition, our entire system, including the image acquisition system, the optical components, and the light patterns has been designed specifically. To test different competing approaches, it is required to adapt this system, while keeping the same environment and components, which is complicated. As a preliminary work, we constructed the DFF and DFD systems to perform comparison in term of accuracy. Nevertheless, during the DFF experiment, we encountered a problem to define sharp regions within the captured images. This is because our test objects have almost textureless (and partial specularities) surface, which cannot be identified by the edge detection algorithms (e.g. Laplacian operator). Therefore, fair comparison using same objects cannot be directly performed, and advanced DFF approaches should be considered. On the other hand, we also faced difficulties in the DFD experiment. The original systems have completely different scales

and magnifications. Therefore, we cannot compare directly with our system unless many optical components are replaced, which will not be appropriate to compare either the qualitative or quantitative results over such systems. To sum up, our system is theoretically prominent to the DFD approach because at least one focus or near-focus image within depth of field exists in the computation. Moreover, it also has less computational extensive comparing to the DFF approach that requires numerous input images.

5.9 Chapter summary

In this chapter, we described the implementation of our depth estimation system. We detailed the specification and characteristic of the experimental setting including the mathematical explanations. The algorithm are in parallel for both the calibrated objects and the test objects. Finally, we reported and discussed on the results obtained from this new prototype. In the next chapter, we summarized the limitation, future work, and perspectives of this approach.

Conclusion

We have introduced a new 3D reconstruction method merging depth from focus and depth from defocus. It can be employed as a stand-alone strategy returning reliable dense depth maps. The method overcomes the problem of weak texture by projecting illumination pattern. Moreover, it does not suffer neither from the correspondence problem nor the occlusion problem found in traditional approaches. Example of applications can be found in the biological specimen analysis, defect metallic component detection, etc.

Limitations and Future Works

Several components in the setup limit the size of the object itself and the maximum change of object depths. However, it is only a matter of scale of the system. This issue can be solved by adjusting smaller or larger optical components for smaller or larger object, respectively, while the algorithm remains unchanged. Therefore, these restrictions are not concerned with the approach methodology.

For further system improvements, we will try to determine the depth for every single screen estimate and run the experiment systematically by robot, to minimize significantly the human error and increase the overall precision of the system. Another future work is to use several mini video projectors and beam splitters to develop a 3D progressive feedback system by controlling the light patterns. Only useful defocused input images are selected iteratively basing on rough 3D model.

The first perspective of this work is to address the problem of depth and reflectance discontinuities of the object (e.g. small and fast variation of the object's depths). The idea is to perform deep analysis of the Gaussian model. To be precise, Gaussian model at these discontinuities will not be close to the reference model (Gaussian distribution). Therefore, these asymmetrical models will be discarded. In order to determine the depth of this problematic region, we plan to employ another type of light pattern within the 3D progressive system. For instance, if this issue occurs from using horizontal stripe pattern, it would not be a problem once we are using vertical stripe pattern. By doing so, we avoid registration problem caused by object movements. Example of scene with depth and reflectance discontinuities are illustrated in figure 5.39.

(a) (b) (c)

Figure 5.39: Scene with depth and reflectance discontinuities

Another perspective is to deal with the scene with multiple objects or with different types of textures variations as exemplified in figure 5.40. The possible solution is to collect sufficient numbers of calibration databases and/or reflectance indexes for all the object materials existing in the scene. The key idea is that some materials provide different reflectance, widths of the blur (CoC), and different spread parameter, accordingly. For instance, some types of plastic have larger blur levels (bigger spread parameters) than the ones observed from metallic materials at the same depth. Therefore, it should be possible to make the order of reflectance and/or blur level with respect to the material types. Then, by matching it with the test objects, we could identify material types and use the current algorithm.

(a) (b) (c)

Figure 5.40: Scene with multiple objects/materials

Our last perspective concerns the estimation of the surface normals by analyzing the deformation of a projected pattern of circles. This elementary projected pattern will be used together with the horizontal and vertical stripe patterns (using different colors projection). The key idea is based on the fact that over smooth depth variations, the circular shapes of the pattern will transform into ellipse shapes.

In order to define the 3D position and orientation of a circle from its projection onto the image plane (an ellipse in the general case), the solution of pose from ellipse-to-circle correspondence described in [95], and [96] is utilized. The solution consists of a series of 3D transformations to the circle and is carried out in two stages: first the orientation of the 3D plane on which the circle lies is determined; and then the center of the circle is computed.

An ellipse in the image plane is of the following form,

$$a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6 = 0. \tag{5.12}$$

We assume the origin is at the principle point, the ellipse defines a cone in 3D becomes

$$a_1x^2 + a_2xy + a_3y^2 + \frac{a_4}{f}xz + \frac{a_5}{f}yz + \frac{a_6}{f}z^2 = 0, \tag{5.13}$$

where f is the focal length of the camera. Then, we follow the procedures explained in [97]. The camera calibration parameters are used to solve the rotation matrix, and eventually return 4 possible solutions of the surface normal characterized by angle θ.

$$\theta_1 = |\theta|$$

$$\theta_2 = -|\theta|$$

$$\theta_3 = \pi + |\theta|$$

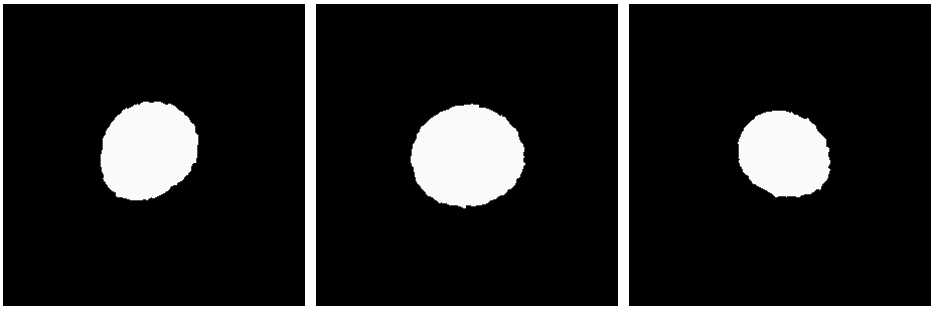
$$\theta_4 = \pi - |\theta|$$

We performed some experiment on the cylindrical object as seen in figure 5.41.

(a) (b)

Figure 5.41: (a) Captured image of cylindrical object (b) with color selection improvement

Three ellipses are extracted and converted into black and white color before computation of surface normal (see figure 5.42).



(a) (b) (c)

Figure 5.42: Extracted ellipse patterns

Because the circle corresponds to a boundary of the physical object it comes from, there must be a direction associated with it and not only an orientation. Therefore, we must also consider the 2 other cases for which the circle is flipped 180 degrees about its diameter. These 4 possible solutions of 3 ellipses are illustrated in figure 5.43 .

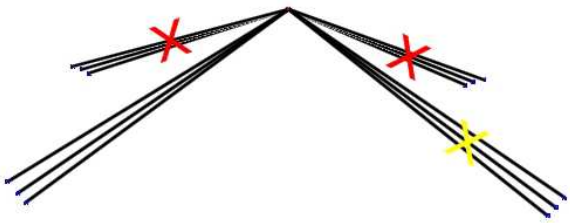


Figure 5.43: Possible solutions of the normal to the object surface

Only one of them is physically correct. Two angles that correspond to the normals with direction toward the object are discarded as shown in red on figure 5.43. We obtain the normal vector N as the following:

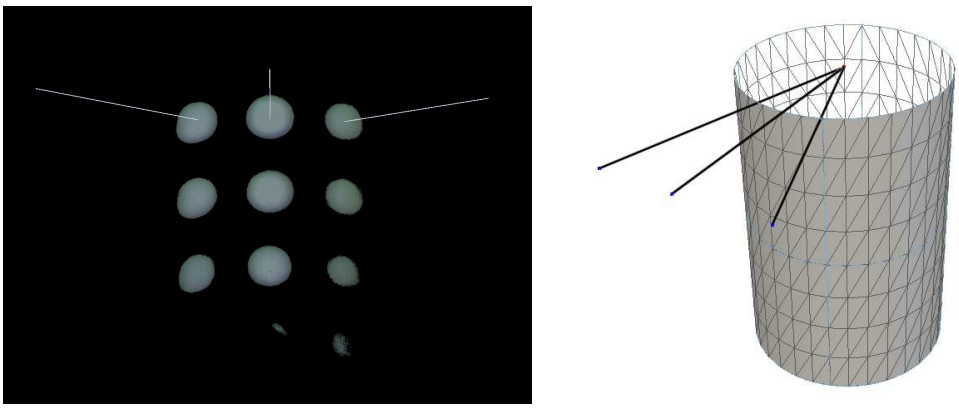
$N_{ellipse1} = (-0.349836, -0.362066, 0.864016)$, or $(0.349835, 0.362063, 0.864017)$

$N_{ellipse2} = (-0.0530441, -0.445498, 0.89371)$, or $(0.0530444, 0.445495, 0.893711)$

$N_{ellipse3} = (-0.273115, 0.425436, 0.862793)$, or $(0.273116, -0.425439, 0.862791)$

Then, to determine between another 2 possibles angles of each ellipse, we require the reconstruction information from vertical and/or horizontal pattern obtained earlier. The correct angle is selected depending on the surrounding depths of this normal position.

Figure 5.44 shows the preliminary result obtained using this approach.



(a) (b)

Figure 5.44: Preliminary results of the surface normal (a) in 2D and (b) in 3D representations

By employing the normal to the surface information, it will help us determine not only the depth but also the direction of the radiance through additional color channel. We expect that this perspective will minimize the error and improve our final 3D reconstruction.

Bibliography

[1] S. Nayar., "Bigshotcamera, computer vision laboratory, columbia university." http://www.bigshotcamera.org/sections/learn/lens/illustrations.html.

[2] R. Szeliski, "3d reconstruction," *Computer Vision*, pp. 505–541, 2011.

[3] Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?," *International Journal of Computer Vision*, vol. 39, no. 2, pp. 141–162, 2000.

[4] A. Sharma, "Projected texture for 3D Object Reconstruction," *PhD. Thesis, International Institute of Information Technology Hyderabad, India*, 2008.

[5] O. Ghita, P. Whelan, and J. Mallon, "Computational approach for depth from defocus," *Journal of Electronic Imaging*, vol. 14, p. 023021, 2005.

[6] M. Watanabe, S. Nayar, and M. Noguchi, "Real-time computation of depth from defocus," in *Proceedings of SPIE the international society for optical engineer*, pp. 14–25, Citeseer, 1996.

[7] M. Kusanagi, K. Terabayashi, K. Umeda, G. Godin, and M. Rioux, "Construction of a 3d model of real-world object using range intensity images," in *Computer and Robot Vision (CRV), 2010 Canadian Conference on*, pp. 317–323, IEEE, 2010.

[8] M. Noguchi and S. Nayar, "Microscopic shape from focus using a projected illumination pattern," *Mathematical and computer modelling*, vol. 24, no. 5-6, pp. 31–48, 1996.

[9] P. Graebling, C. Boucher, C. Daul, and E. Hirsch, "3D sculptured surface analysis using a structured-light approach," in *Proceedings of SPIE*, vol. 2598, p. 128, 1995.

[10] P. Besl, "Active, optical range imaging sensors," *Machine vision and applications*, vol. 1, no. 2, pp. 127–152, 1988.

[11] B. Curless and M. Levoy, "Better optical triangulation through spacetime analysis," p. 987, 1995.

[12] H. Feng, Y. Liu, and F. Xi, "Analysis of digitizing errors of a laser scanning system," *Precision Engineering*, vol. 25, no. 3, pp. 185–191, 2001.

[13] J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, no. 4, pp. 827–849, 2004.

[14] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, and R. Scopigno, "A low cost 3d scanner based on structured light," vol. 20, no. 3, pp. 299–308, 2001.

[15] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 150–162, 1994.

[16] S. Tosovic, R. Sablatnig, and M. Kampel, "On combining shape from silhouette and shape from structured light," vol. 108, p. 118, 2002.

[17] A. Chiuso and S. Soatto, "Mfm: 3-d motion from 2-d motion causally integrated over time-part i: Theory," 2000.

[18] Y. Ohta and T. Kanade, "Stereo by intra-and inter-scanline search using dynamic programming," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 139–154, 1985.

[19] I. Cox, S. Hingorani, S. Rao, and B. Maggs, "A maximum likelihood stereo algorithm," *Computer vision and image understanding*, vol. 63, no. 3, pp. 542–567, 1996.

[20] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on pattern analysis and machine intelligence*, pp. 1222–1239, 2001.

[21] J. Sun, N. Zheng, and H. Shum, "Stereo matching using belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 787–800, 2003.

[22] S. Barnard, "Stochastic stereo matching over scale," *International Journal of Computer Vision*, vol. 3, no. 1, pp. 17–32, 1989.

[23] Y. Furukawa and J. Ponce, "Accurate camera calibration from multi-view stereo and bundle adjustment," *International journal of computer vision*, vol. 84, no. 3, pp. 257–268, 2009.

[24] V. Kolmogorov and R. Zabih, "Multi-camera scene reconstruction via graph cuts," *Computer Vision-ECCV 2002*, pp. 8–40, 2002.

[25] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," 2006.

[26] K. Ikeuchi and B. Horn, "Numerical shape from shading and occluding boundaries," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 141–184, 1981.

[27] R. Szeliski, "Fast shape from shading," *CVGIP: Image Understanding*, vol. 53, no. 2, pp. 129–153, 1991.

[28] O. Vaga and Y. Yang, "Shading logic: A heuristic approach to recover shape from shading," *IEEE transactions on pattern analysis and machine intelligence*, pp. 592–597, 1993.

[29] B. Horn, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," 1970.

[30] J. Oliensis and P. Dupuis, "A global algorithm for shape from shading," pp. 692–701, 1993.

[31] M. Bichsel and A. Pentland, "A simple algorithm for shape from shading," pp. 459–465, 1992.

[32] A. Pentland, "Local shading analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 2, pp. 170–187, 1984.

[33] C. Lee and A. Rosenfeld, "Improved methods of estimating shape from shading using the light source coordinate system," *Artificial Intelligence*, vol. 26, no. 2, pp. 125–143, 1985.

[34] A. Pentland, "Shape information from shading: a theory about human perception," *Spatial Vision, 4*, vol. 2, no. 3, pp. 165–182, 1989.

[35] P. Tsai and M. Shah, "Shape from shading using linear approximation," *Image and Vision Computing*, vol. 12, no. 8, pp. 487–498, 1994.

[36] D. Weinshall and C. Tomasi, "Linear and incremental acquisition of invariant shape models from image sequences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 5, pp. 512–517, 1995.

[37] T. Morita and T. Kanade, "A sequential factorization method for recovering shape and motion from image streams," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 8, pp. 858–867, 1997.

[38] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion," *Computer Vision-ECCV*, pp. 709–720, 1996.

[39] M. Irani and P. Anandan, "Factorization with uncertainty," *Computer Vision-ECCV*, pp. 539–553, 2000.

[40] J. Costeira and T. Kanade, "A multibody factorization method for independently moving objects," *International Journal of Computer Vision*, vol. 29, no. 3, pp. 159–179, 1998.

[41] J. Yan and M. Pollefeys, "A factorization-based approach to articulated motion recovery," vol. 2, pp. 815–821, 2005.

[42] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3d shape from image streams," vol. 2, pp. 690–696, 2000.

[43] W. Brand, "Morphable 3d models from video," vol. 2, pp. II–456, 2001.

[44] D. W. Jacobs, "Linear fitting with missing data for structure-from-motion," *Computer Vision and Image Understanding*, vol. 82, pp. 57–81, 2001.

[45] Y. Ohta, K. Maenobu, and T. Sakai, "Obtaining surface orientation from texels under perspective projection," pp. 746–751, 1981.

[46] D. Blostein and N. Ahuja, "Shape from texture: Integrating texture-element extraction and surface estimation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 12, pp. 1233–1251, 1989.

[47] R. Bajcsy and L. Lieberman, "Texture gradient as a depth cue," *Computer Graphics and Image Processing*, vol. 5, no. 1, pp. 52–67, 1976.

[48] L. Brown and H. Shvaytser, "Surface orientation from projective foreshortening of isotropic texture autocorrelation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 6, pp. 584–588, 1990.

[49] J. Gårding, "Shape from texture for smooth curved surfaces in perspective projection," *Journal of Mathematical Imaging and Vision*, vol. 2, no. 4, pp. 327–350, 1992.

[50] M. Clerc and S. Mallat, "The texture gradient equation for recovering shape from texture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 536–549, 2002.

[51] D. Forsyth, "Shape from texture without boundaries," *Computer Vision-ECCV 2002*, pp. 43–66, 2002.

[52] J. Ens and P. Lawrence, "An investigation of methods for determining depth from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 2, pp. 97–108, 1993.

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

[54] I. Lertrusdachakul, Y. Fougerolle, and O. Laligant, "A novel 3D reconstruction approach by dynamic (de) focused light," in *Proceedings of SPIE*, vol. 7538, p. 75380L, 2010.

[55] T. Choi, M. Asif, and J. Yun, "Three-dimensional shape recovery from focused image surface," vol. 6, pp. 3269–3272, 1999.

[56] T. Darrell and K. Wohn, "Pyramid based depth from focus," pp. 504–509.

[57] B. Horn and E. Weldon, "Direct methods for recovering motion," *International Journal of Computer Vision*, vol. 2, no. 1, pp. 51–76, 1988.

[58] J. Tenenbaum, "Accommodation in computer vision.," tech. rep., DTIC Document, 1970.

[59] R. Jarvis, "Focus optimization criteria for computer image processing," *Microscope*, vol. 24, no. 2, pp. 163–180, 1976.

[60] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques (journal paper)," *Optical Engineering*, vol. 32, no. 11, pp. 2824–2836, 1993.

[61] Y. Xiong and S. Shafer, "Depth from focusing and defocusing," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 68–68, Citeseer, 1993.

[62] N. Xu, K. Tan, H. Arora, and N. Ahuja, "Generating omnifocus images using graph cuts and a new focus measure," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 4, pp. 697–700, IEEE.

[63] A. Rajagopalan and S. Chaudhuri, "A variational approach to recovering depth from defocused images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1158–1164, 1997.

[64] M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.

[65] P. Favaro, A. Duci, and G. Rostock, "A theory of defocus via Fourier analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[66] S. Shim, A. Malik, and T. Choi, "Accurate shape from focus based on focus adjustment in optical microscopy," *Microscopy Research and Technique*, vol. 72, no. 5, pp. 362–370, 2009.

[67] S. Li, *Markov random field modeling in image analysis*. Springer-Verlag New York Inc, 2009.

[68] G. Surya and M. Subbarao, "Depth from defocus by changing camera aperture: A spatial domain approach," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pp. 61–67, IEEE, 1993.

[69] M. Watanabe and S. Nayar, "Rational filters for passive depth from defocus," *International Journal of Computer Vision*, vol. 27, no. 3, pp. 203–225, 1998.

[70] M. Gokstorp, "Computing depth from out-of-focus blur using a local frequency representation," in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1, pp. 153–158, IEEE, 1994.

[71] A. Mennucci and S. Soatto, "On observing shape from defocused images," in *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pp. 550–555, IEEE, 1999.

[72] S. Soatto and P. Favaro, "A geometric approach to blind deconvolution with application to shape from defocus," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 2, pp. 10–17, IEEE, 2000.

[73] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 406–417, 2005.

[74] P. Favaro, S. Soatto, M. Burger, and S. Osher, "Shape from defocus via diffusion," *IEEE transactions on pattern analysis and machine intelligence*, pp. 518–531, 2008.

[75] C. Zhou, S. Lin, and S. Nayar, "Coded aperture pairs for depth from defocus," in *Proc. International Conference on Computer Vision, Kyoto, Japan*, 2009.

[76] F. Moreno-Noguer, P. Belhumeur, and S. Nayar, "Active refocusing of images and videos," in *ACM SIGGRAPH 2007 papers*, pp. 67–es, ACM, 2007.

[77] C. Hermans, Y. Francken, T. Cuypers, and P. Bekaert, "Depth from sliding projections," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1865–1872, IEEE, 2009.

[78] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, 2011.

[79] A. Savakis and H. Trussell, "On the accuracy of psf representation in image restoration," *Image Processing, IEEE Transactions on*, vol. 2, no. 2, pp. 252–259, 1993.

[80] M. Born and E. Wolf, "Principles of optics, 7-th ed," *Cambridge U. Press, Cambridge, UK*, 1999.

[81] W. Smith, *Modern lens design: A resource manual*. McGraw-Hill (New York), 1992.

[82] G. Arfken, H. Weber, and H. Weber, *Mathematical methods for physicists*, vol. 148. Academic press New York, 1995.

[83] E. Hecht, "Optics 4th edition," *Optics 4th edition by Eugene Hecht Reading, MA: Addison-Wesley Publishing Company, 2001*, vol. 1, 2001.

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

[85] N. Asada, H. Fujiwara, and T. Matsuyama, "Edge and depth from focus," *International Journal of Computer Vision*, vol. 26, no. 2, pp. 153–163, 1998.

[86] S. Das and N. Ahuja, "Performance analysis of stereo, vergence, and focus as depth cues for active vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 12, pp. 1213–1219, 1995.

[87] A. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 9, no. 4, pp. 523–531, 1987.

[88] M. Subbarao and N. Gurumoorthy, "Depth recovery from blurred edges," in *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR'88., Computer Society Conference on*, pp. 498–503, IEEE, 1988.

[89] A. Pentland, S. Scherock, T. Darrell, and B. Girod, "Simple range cameras based on focal error," *JOSA A*, vol. 11, no. 11, pp. 2925–2934, 1994.

[90] S. Nayar, M. Watanabe, and M. Noguchi, "Real-time focus range sensor," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 12, pp. 1186–1198, 1996.

[91] L. Zhang and S. Nayar, "Projection defocus analysis for scene capture and image display," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 907–915, 2006.

[92] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[93] E. Billauer, "peakdet: Peak detection using matlab," *Online, http://www. billauer. co. il/peakdet. htm.*

[94] M. Hanke, "A regularizing levenberg-marquardt scheme, with applications to inverse groundwater filtration problems," *Inverse problems*, vol. 13, p. 79, 1997.

[95] D. Forsyth, J. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell, "Invariant descriptors for 3d object recognition and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 971–991, 1991.

[96] M. Dhome, J. Lapreste, G. Rives, and M. Richetin, "Spatial localization of modelled objects of revolution in monocular perspective vision," *Computer Vision ECCV 90*, pp. 475–485, 1990.

[97] M. Costa and L. Shapiro, "3d object recognition and pose with relational indexing* 1," *Computer Vision and Image Understanding*, vol. 79, no. 3, pp. 364–407, 2000.

[98] A. Pentland, T. Darrell, M. Turk, and W. Huang, "A simple, real-time range camera," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 256–261, 1989.

[99] M. Hanke, "The regularizing levenberg-marquardt scheme is of optimal order," *Journal of Integral Equations and Applications*, vol. 22, no. 2, pp. 259–283, 2010.

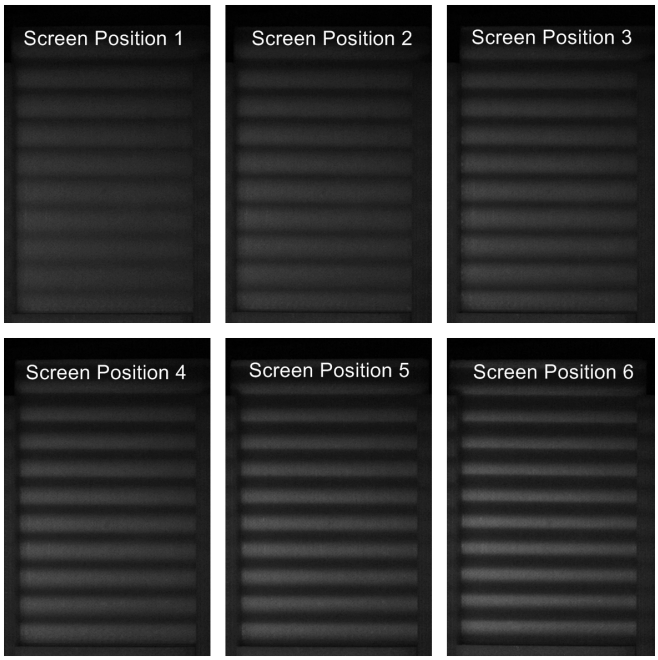
[100] J. Gibson, "The perception of visual surfaces," *The American Journal of Psychology*, vol. 63, no. 3, pp. 367–384, 1950.

List of Figures

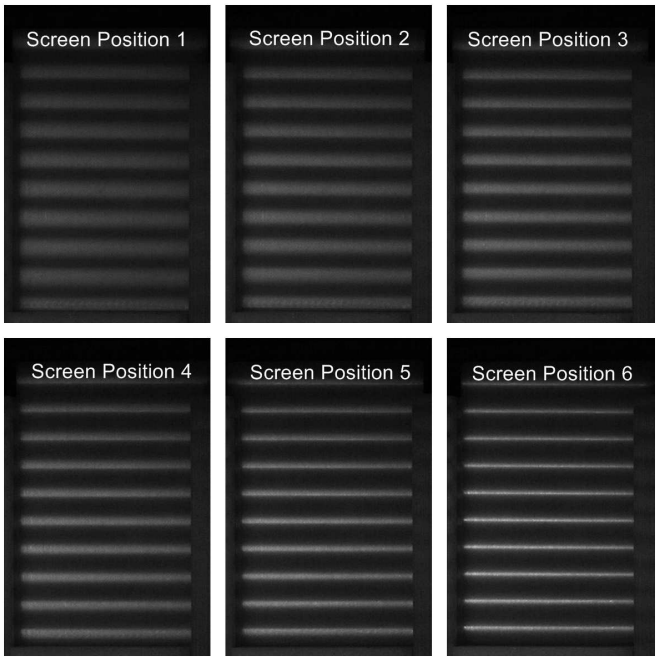
List of Tables

Appendix A

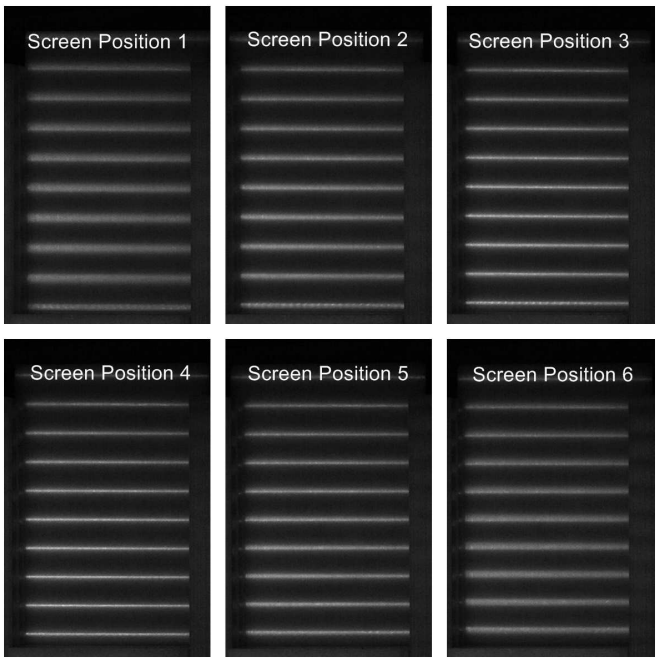
Additional Images

Calibrated Object	Captured images from different screen positions
Depth D1	

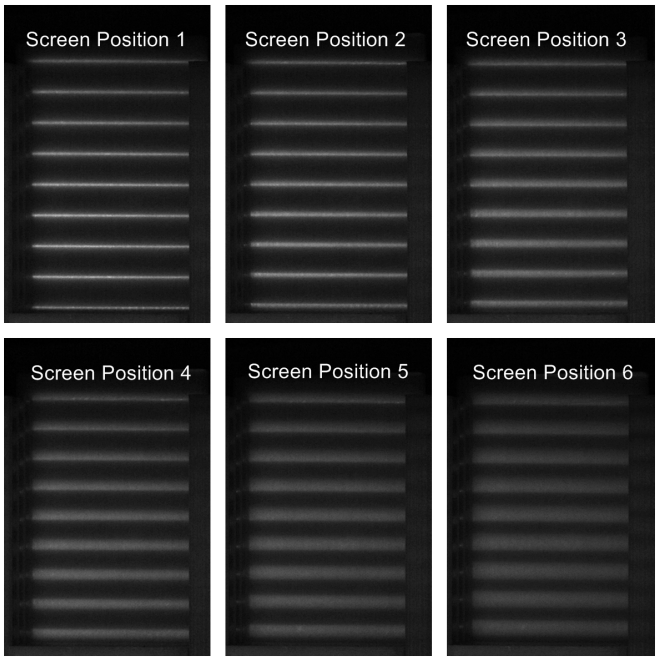
Depth D2



Depth D3



Depth D4



Depth D5

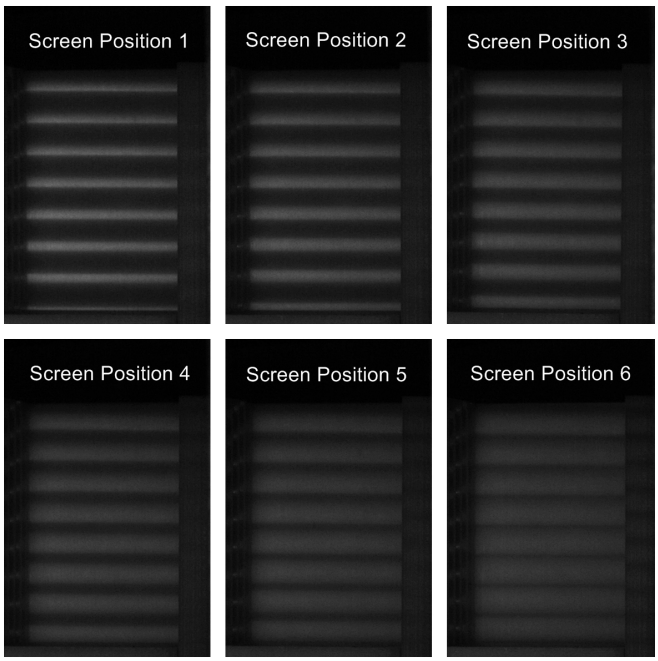


Table A.1: Captured images of the calibrated objects

Appendix B

List of publications

Journal

[Lertrusdachakul et al., 2011] Lertrusdachakul, I., Fougerolle, Y. D., and Laligant, O. (2011). Dynamic (de)focused Projection for 3D Reconstruction. *Optical Engineering, 50(11).*

International Conferences

[Lertrusdachakul et al., 2010a] Lertrusdachakul, I., Fougerolle, Y. D., and Laligant, O. (2010a). A Novel 3D Reconstruction Approach by Dynamic (de)focused Light. *Electronic Imaging - Image processing : Machine vision applications III, Proceedings of the SPIE, 7538.*

[Lertrusdachakul et al., 2010b] Lertrusdachakul, I., Fougerolle, Y. D., and Laligant, O. (2010a). Depth from dynamic (de)focused projection. *International Conference of Image and Vision Computing New Zealand (IVCNZ 2010), Proceedings of the IEEE Explore, 978-1-4244-9631-0.*