天道酬勤

— 尚书

University of Alberta

3D Reconstruction of Transparent and Specular Objects

by

Ding Liu

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Computing Science

Fall 2013
Edmonton, Alberta

This thesis is dedicated to my dearest mom and dad.

Abstract

3D reconstruction of transparent and specular objects is a very challenging topic in computer vision. The goal is to get the 3D information of the points on the surface of a transparent or specular object and accumulate the points to form the reconstructed surface. For opaque objects, the structured light methods can be used with good results. For transparent and specular objects, which have complex interior and exterior structures that can reflect and refract light in a complex fashion, it is difficult, if not impossible, to use the traditional structured light methods to do the reconstruction.

In this thesis, a frequency-based 3D reconstruction method based on the frequency-based matting method is introduced. Similar to the structured light methods, a set of frequency-based patterns are projected to the object, and a camera captures the scene at the same time. Each pixel of a captured image is analyzed along the time axis and the signal is transformed to the frequency-domain using the Discrete Fourier Transformation. Since the frequency is only determined by the source that creates it, the frequency of the signal can uniquely identify the location of the pixel in the patterns. In this way, the correspondences between the pixels in the captured images and the points in the patterns can be acquired. Using a new labelling procedure developed in this research, the surface of transparent and specular objects can be reconstructed with very encouraging results.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Herbert Yang, for his constant help, guidance and outstanding supervision. I want to thank him for not giving me up and always believing in me. His excellent insight in academics gave me valuable directions of my research. His dedication to a quality research and work provided major source of inspiration to me. Second, I would like to thank my committee members, Dr. Janelle Harms and Dr. Anup Basu, for taking their time to read my thesis and participate in my defense. Third, I also would like to thank my committee chair, Dr. Abram Hindle, for taking his time to be the chair of my defense.

I would like to express my thanks and appreciation to all the members from our Computer Graphics Group - Xida Chen, Rouzbeh Maani, Yi Gu, Timothy Yau, Samaneh Eskandari, Jian Li and Mohsen Taghaddosi. I want to thank them for their good advice and helpful tips during our group meetings and discussions. Especially, I would like to thank Xida Chen, for helping me to learn basic concepts about 3D reconstruction and providing helpful advice.

I would like to acknowledge the Department of Computing Science at University of Alberta for providing me the great opportunity to pursue my master degree. Financial support from NSERC and the Department of Computing Science is gratefully acknowledged.

Last but not least, I would like to express my undescribable gratitude and deepest love to my parents for their understanding and unconditional support. This thesis is dedicated to my beloved parents.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

3D reconstruction is the procedure of capturing the shape or surface structure of objects. The goal is to acquire the 3D information of each point on the surface of an object, 3D coordinates, depth and normal. By assembling all these points, the surface of the object can be reconstructed.

In the past few decades, 3D reconstruction has been widely used in robotics, game design, movies and cartoon design, interior decoration, automation and so on. For example, in movie-making, it is difficult, if not impossible, to use existing software to draw the delicate shape of an object manually. Instead, we can use 3D reconstruction methods to get the structure of that object first and then make some appropriate adjustments or directly use it. Another example would be in digitization. Nowadays, more and more objects are required to be "stored" in a digital format in a database for ease of access over the Internet or for ease of storage. To digitize the shape and structure of these objects, many 3D reconstruction methods have been developed.

There are two main types of methods to do 3D reconstruction, which are active methods and passive methods. Active methods use devices such as projectors, monitors or laser emitters to cast patterns onto objects, and use receivers, such as camera(s), to simultaneously capture the scene. By analyzing the captured images with information of the relative positions of the devices, the 3D information of the points can be obtained. Passive methods usually use only camera(s) to capture the scene from different angles and use images as their only input to do reconstruction. Since passive methods tend to have less accurate results than active methods, nowadays, active methods have drawn more and more attentions.

In the realm of active methods, techniques using structured light are the most commonly used ones. Structured light methods cast a set of coded patterns onto the object and use a camera to capture the scene simultaneously. Hence, the correspondence between a pixel in the projector image and the corresponding pixel in the captured image can be easily established. When the object is opaque and has a Lambertian surface, which means that the surface can reflect the incoming light uniformly, it is easy to find corresponding points between the pattern and the captured image. The camera center, the projector center and the point on the surface form a triangle, as shown in Fig. 1.1. The procedure of triangulation is to solve this triangle and find the 3D information of the point X. If the camera and the projector are calibrated, the position of the camera center and the projector center are known. The direction from the camera center to the corresponding pixel can be obtained. As well, the direction from the projector center to the point on the pattern can also be calculated. Hence, the intersection of these two directions, which is the point X on the object surface, can be acquired.

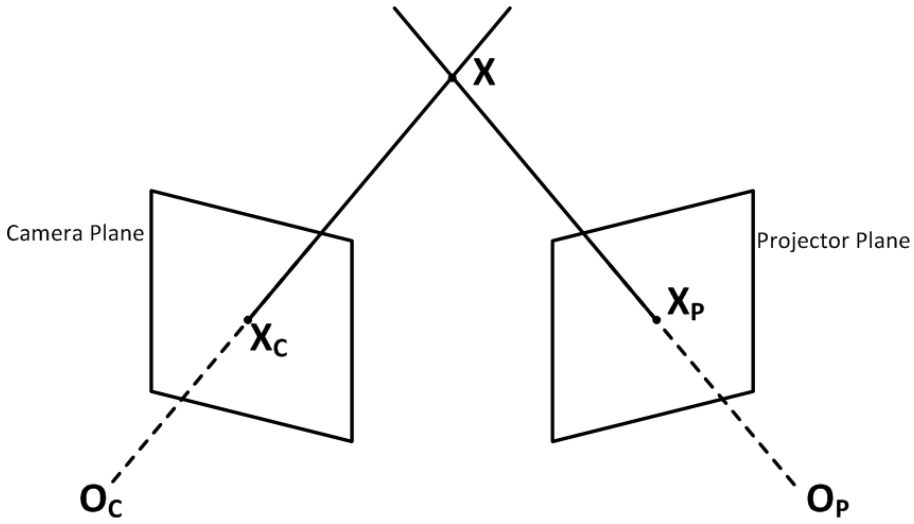


Figure 1.1: An illustration of triangulation. O_C is the camera center. O_P is the projector center. X_C and X_P are the corresponding points, respectively, on the camera plane and the projector plane. X is the intersection point on the surface of the object.

As we can see, the key piece of information for triangulation is to correctly find the correspondences. For opaque objects with Lambertian surface, it is quite easy and straightforward. Many methods [5, 6, 7, 8] can reconstruct well-defined surface using structured light methods. However, for objects that have poor reflection or anisotropic properties(Fig. 1.2a and Fig. 1.2b), which means that the reflection is

either weak or non-uniform, 3D reconstruction is still an active research topic.

(a) A transparent trophy with weak reflection

(b) A metal cup with anisotropic properties

(c) A vase with a complex surface structure

Figure 1.2: Examples of objects that are difficult to be reconstructed using existing methods.

Different from Lambertian surface, specular surface reflects light with a dominant angle, causing a strong specular highlight. The specular highlight can make it difficult, if not impossible, to acquire the texture of the highlight point. Hence, it is difficult to reconstruct the objects with specular surface. Additionally, the specular highlight is view dependent and creates a problem for passive stereo methods.

Among all types of objects, the most difficult type to do 3D reconstruction is transparent objects. Transparent objects are very common and can be made of many different types of materials from crystal, quartz to water. It is difficult to reconstruct them because they are *optically active*, which means that they interact with light in a complex fashion [4]. For example, Fig. 1.2c shows a vase, which has a complex surface structure that not only can reflect but also refract light, which results in highlight and caustic effects. Hence, when we use traditional 3D reconstruction methods, either active or passive methods, the optical effects will lead to erroneous results.

Not only transparent objects with complex surface are challenging to be reconstructed, even objects that are transparent and with a smooth surface are also optically active and hard to be reconstructed. Sometimes even a human eye cannot discern the structure, let alone using cameras and images. Indeed, eyeglasses are made for people to "see through" without having the structure of the eyeglasses observed. People may walk into a glass door just because they did not see the reflection on it and everything on the other side was clear and without distortion. Another example would be the trophy as shown in Fig. 1.2a, which is transparent. When we want to use a traditional method, for example, using a projector to cast

Gray code patterns onto it, most light would transmit through the object and get reflected by the background wall. In that case, the reflected light will interfere with the real reflection from the object surface, making it difficult to use reflection to do 3D reconstruction.

Some existing methods [9, 10] attempt to address some of the problems related to the transparent properties, but they all have certain shortcomings and limitations. For example, Trifonov et al. [11] needs to suspend the object in a poisonous and caustic solution, which is not only dangerous to the researchers doing the experiments but also may damage the object. Methods by [12, 13] can only be applied well to a small group of objects that have a simple shape, with a known refraction index. Yeung et al. [14] needs user interactions during the process. From the viewpoint of setup complexity, methods by [15, 16] use too many devices in their experiments. Morris et al. [17] needs to move their setup during experiments, leading to more errors.

Comparing the limitations and shortcomings stated above, the structured light methods have a simple and stable setup, and the experiment is non-contact with the object. Most importantly, they can be applied to a wider range of objects. Since structured light methods use only reflection, any object that can reflect light can be reconstructed by structured light methods in theory. However, as discussed before, the real challenge is that for transparent and specular objects, the reflection is either weak or non-uniform. Hence, it is difficult to find the correct correspondences between points on the pattern and pixels on the image.

Although in 3D reconstruction of transparent and specular objects, finding the correspondences is very challenging, researchers have made progress in another closely related research topic called environment matting, whose main challenge is also about finding the correct correspondences. An environment matte of a transparent object shows how this object refracts and reflects light from the environment. Since the foreground object is transparent, it can converge and disperse the environment light. Hence, the main issue of environment matting is to distinguish different components and to find the correspondence map of each pixel in the captured image to the pixels of background image.

One of the elegant solutions to environment matting is the frequency-based environment matting method proposed by Zhu and Yang [4]. The method was inspired

by the fact that a time domain signal has a unique decomposition in the frequency domain. They use that idea to successfully find the correct correspondences between the backdrop patterns and the obtained images.

To my knowledge, none of the existing methods have incorporated the environment matting methods with 3D reconstruction methods for the purpose of finding the correct correspondences.

1.2 Contributions

The goal of this work is to use the structured light methods incorporated with the environment matting methods to do 3D reconstruction of transparent and specular objects. Based on the challenges stated above, the contributions that are achieved in this thesis are identified as follows:

First, the proposed method incorporates the 3D reconstruction method with the environment matting method, and can find correct correspondences for transparent and specular objects between points on projected patterns and pixels on captured images.

Second, a new labelling method is proposed to successfully find the correct points on the surface of the object.

Third, the proposed method is applicable to both transparent and specular objects. For transparent objects, this method obtains accurate and robust results for objects that have complex structures as well as objects that are totally transparent that have challenged other methods. For specular objects, this method can also acquire accurate and robust results for objects that have anisotropic properties.

1.3 Organization of the thesis

The rest of the thesis is organized as follows.

In Chapter 2, related methods for 3D reconstruction of opaque objects, transparent objects, and specular objects are presented and reviewed. Relevant methods for environment matting are introduced. As well, concepts and background related to the proposed method are discussed.

In Chapter 3, the frequency-based method for environment matting to obtain the correspondences between images taken by a camera and patterns casted by a

projector is introduced. A new labelling method to get the point of reflection, and a triangulation method to get the 3D points on the surface of the object are presented. Some post-processing methods are then discussed.

In Chapter 4, details of the experiments and results are presented. Experiments with 7 objects are conducted. The results are compared with the ground truth as well as with results using a classic method [18].

Chapter 5 gives the conclusions and a road-map for future work.

Chapter 2

Background and Related Works

3D reconstruction was originally introduced to acquire the shape and surface structure of opaque objects with diffuse surface (Fig. 2.1). One of the representative techniques is the structured light methods. These methods can achieve accurate results when the objects are opaque and have a diffuse surface, but may fail when the objects are transparent or have a specular surface (Fig. 2.1). The reason is because for transparent and specular objects, the reflection is either weak or non-uniform, and sometimes may be interfered by other light effects, making it difficult to find the correct correspondences between the points on the projected patterns and the pixels in the captured images. Since the correspondences are the key information to do accurate triangulation, inaccurate correspondences lead to wrong reconstruction results.

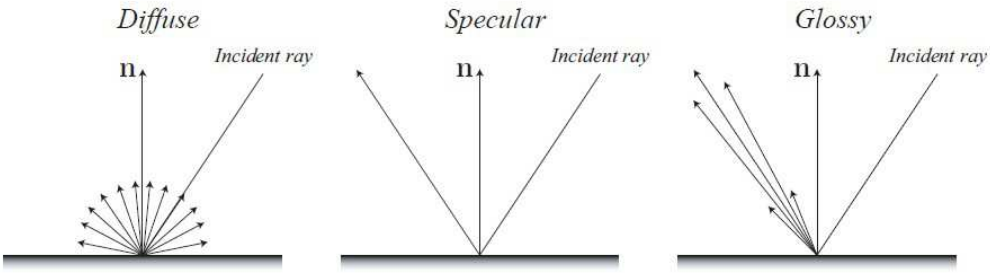


Figure 2.1: Illustrations of the different types of reflectance surfaces [1].

3D reconstruction of transparent and specular objects has been an active research topic for many years because it is a challenging task and has a wide range of applications [19]. Researchers have devoted years to study this problem, aiming to come up with a method that can efficiently acquire accurate 3D information of points on the surface of the transparent and specular objects, but only with limited success. In this chapter, representative methods for 3D reconstruction of transparent

and specular objects are presented and reviewed.

When using a structured light method for transparent and specular objects, the biggest challenge is to find the correct correspondences, and no existing method has perfectly solved this problem. However, methods to do environment matting have solved the correspondence problem to some extent. Since these methods can be adapted for 3D reconstruction, in this chapter, related methods for environment matting are also reviewed.

2.1 3D reconstruction of opaque objects

2.1.1 Passive methods versus active methods

Passive methods normally use information such as colour, edges, corners, textures, or higher level descriptors, such as SIFT [20, 21, 22, 23], SURF [24], or DAISY [25], to find correspondences among different images. Since the cameras from different viewing angles have been calibrated, the triangulation can be done similar to that shown in Fig. 1.1. One advantage of passive stereo methods is that they only need camera(s) and no other light emitting devices. By moving the camera around an object, the 3D reconstruction of an object can be obtained. However, when the surface of the object is not sufficiently textured, it is difficult to find features, and hence, correspondence. In this case, passive methods may fail in finding the correct corresponding points. In particular, for the surface of specular or transparent objects, the appearance of the surface depends on the environment as well as on the viewing angle. Hence, the correspondences will not be correct. For objects with a uniform texture, it is difficult, if not impossible, to detect the corresponding features from different views. Hence, finding the correct correspondences is very challenging and passive methods often fail to do accurate 3D reconstruction in these cases.

For active methods, devices such as projectors, monitors and laser emitters are needed in addition to camera(s). Rather than passively capturing the image of an object, active methods cast specially designed coded patterns or points onto the object, and use camera(s) to capture the scene with the projected patterns. Using coded patterns, each projected point on the surface of an object is unique and can be identified. Hence, the correspondences between the pixels on the captured images and the pixels on the projected patterns can be easily established. Using the triangulation method shown in Fig. 1.1, the 3D information of the point on the

surface of the object can be calculated.

The difference between passive methods versus active methods is shown in Table 2.1.

Table 2.1: Passive methods versus active methods

	Passive methods	Active methods
Setup	Only camera(s)	Camera(s) and other light emitting devices, such as projectors, monitors or laser emitters
Information of input	Images	Images and information of the patterns

The focus of this thesis is to use the structured light and coded patterns to do 3D reconstruction of static transparent and specular objects. Hence, representative structured light methods for opaque objects are reviewed below.

2.1.2 3D reconstruction using structured light

Fig. 2.2 is an illustration of structured light. The projector casts a pattern onto an opaque object. Since the object has a Lambertian surface, the pattern gets reflected uniformly. The camera, which is located on the same side as the projector, captures the scene with the projected pattern.

There are three steps in 3D reconstruction using structured light. First, the projector emits each coded pattern sequentially onto the object, with the camera capturing the image synchronously. Second, the captured images are analyzed and the correspondence between the point on the pattern and the pixel on the captured image are established. Third, the triangulation is conducted and the depth information of the point on the surface of the object is obtained. To do triangulation, see Fig. 1.1, XX_C can be computed based on the camera center to the pixel on the camera plane and X_PX can be obtained from the projector center to the corresponding point on the projector plane. Ideally, these two straight lines "intersect" at X, which is the reconstructed point on the surface of the object. In practice, because of measurement errors and noise, these two straight lines do not intersect, in which case, the mid-point of their common perpendicular line segment is treated as the intersection.

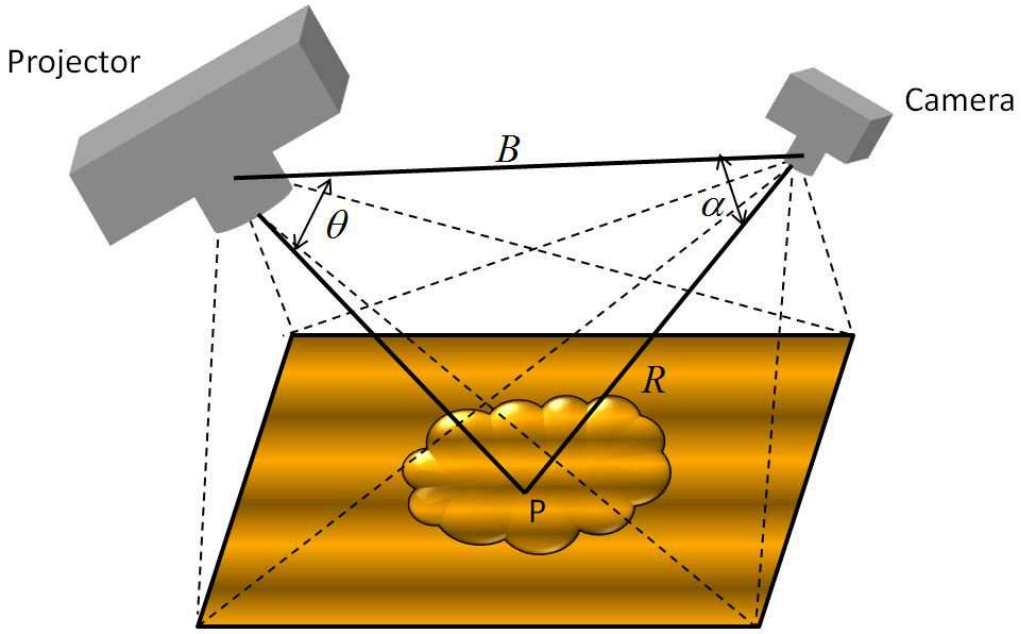


Figure 2.2: An illustration of structured light (redrawn based on [2]).

Eq. 2.1 [2] shows the basic idea of the geometric relationship between the knowns and the unknowns.

$$R = B\frac{sin(\theta)}{sin(\alpha + \theta)} \tag{2.1}$$

where, see Fig. 2.2, R denotes the distance between the camera center and point P on the surface of an object. B is the base line, showing the distance between the camera center and the projector center. θ is the angle of the direction from the projector center to point P with the base line. Similarly, α is the angle of the direction from the camera center to point P with the base line. The camera center, the projector center, and point P on the surface of the object form a triangle. With calibration, the position of the projector center can be easily calculated in the coordinate system with the camera center as the origin. After finding the corresponding points, the directions of the incoming light and the outgoing light can be obtained, and the values of θ and α are known. With these values, Eq. 2.1 can be solved and the value of R is computed. Since the direction of the incoming light and the distance R are obtained, the position of P can be obtained, where P is the reflection point on the surface of the object. With this method, all the points on the surface of the object can be calculated, and the whole object can be reconstructed by assembling these points.

As stated by Geng [2], the structured light methods can be classified into two groups. One group [18, 26, 27] is the sequential projection techniques, which use a sequence of images to do the reconstruction. Normally, they can only be used when the object is static, because multiple coded patterns are projected onto the object and are decoded in order to get the correspondences. The other group [28, 29, 30] is the single image techniques, which can be used for moving objects. Normally, methods using multiple images are more accurate than those using a single image.

In the experiments using the proposed method in this thesis, all of the objects are motionless, and the patterns used are the sequential ones. Hence, only related sequential projection methods are reviewed below.

2.1.3 Sequential projection methods

One of the most fundamental sequential projection methods is the method using the Gray code patterns. Shown in Fig. 2.3 are the Gray code patterns, which is a kind of binary patterns. Typically two groups of patterns, the normal patterns and their complement or commonly known as the inverse patterns, are designed for projection. The goal of the inverse patterns is to suppress the effect of noise. Instead of using a threshold to decide if a projected pixel is a 1 or a 0, when a pattern is projected onto an object, its inverse is also projected. For each pixel, a decision is made based on whether or not the intensity of this pixel in the normal image is higher than that in the inverse image. In this way, a sequence of digits, such as "0110001101", is obtained for each pixel, and each sequence can uniquely identify the one-dimensional position in the pattern. In order to uniquely identify the two-dimensional position in the pattern, usually the patterns are designed in two orientations, vertical and horizontal. For a projector with a resolution of $n \times m$, a number of $log_2(n \times m)$ patterns are needed. For example, for a projector with a resolution of 1024×768, 10 vertical patterns and 10 horizontal patterns are designed to identify each pixel of the patterns. With the correspondences between pixels on the captured images and pixels on the projected patterns, the triangulation can be done, and the 3D information of the points on the surface of the object can be obtained.

There are advantages and disadvantages of using the Gray code patterns. The advantage is its resilience to errors and noise, since it is only needed to make a "true

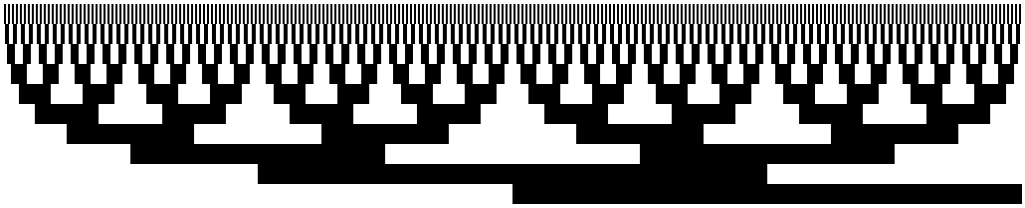


Figure 2.3: Gray Code Patterns.

or false" decision. However, if a higher resolution is required, the number of the patterns has to be increased, which is both time and space consuming. In addition, since a typical camera has a much higher resolution than that of a projector, when a pattern is projected onto an object, one projected pixel covers more than one pixel in the captured image. Hence, interpolation is often required to compute the corresponding point in the captured image. Moreover, the Gray code patterns tend to fail in situations of inter-reflections, subsurface scattering, and defocus [8]. To resolve these problems, researchers [31, 8, 18] have proposed extensions of the Gray code projection.

Valkenburg et al. [18] include lens distortion, substripe estimation and subpixel estimation into the models for camera and projector, and achieve a good performance in the order of 0.3 mm [18]. Tsai et al. [5] combine the Gray code patterns with the sub-pixel technology to increase the accuracy of the results. Aliaga et al. [32] present a self-calibrating photogeometric reconstruction method. Since the projector and cameras do not need to be calibrated, they can change their positions during experiments. Gupta et al. [8, 33] use simple logical operations to increase the resilience to errors caused by global illumination, such as inter-reflection, subsurface scattering, and defocus.

Apart from the Gray code patterns, phase shifting [34, 35, 36, 37] is also a set of coding methods that is commonly used by researchers. The patterns are usually presented by Eq. 2.2 [2]

$$I_1(x,y) = I_0(x,y) + I_{mod}(x,y)cos(\phi(x,y) - \theta) \tag{2.2a}$$

$$I_2(x,y) = I_0(x,y) + I_{mod}(x,y)cos(\phi(x,y)) \tag{2.2b}$$

$$I_3(x,y) = I_0(x,y) + I_{mod}(x,y)cos(\phi(x,y) + \theta) \tag{2.2c}$$

where $I_1(x,y)$, $I_2(x,y)$, and $I_3(x,y)$ are the intensities of the coded patterns, $I_0(x,y)$ the direct component (background), $I_{mod}(x,y)$ the modulation signal amplitude,

$\phi(x, y)$ the phase, and θ the constant phase-shift angle [2]. When θ is equal to $\frac{2}{3}\pi$, Eq. 2.2 can be solved and $\phi(x, y)$ can be obtained,

$$\phi(x, y) = tan^{-1}\left(\sqrt{3}\frac{I_0(x, y) - I_2(x, y)}{2I_1(x, y) - I_0(x, y) - I_2(x, y)}\right). \tag{2.3}$$

Since there is an arctangent calculation, the value of $\phi(x, y)$ is limited within $(-\pi+2k\pi, \pi+2k\pi)$, where k is an integer representing the projection period. Because of the different values of k, the values of $\phi(x, y)$ are discontinuous. The procedure to solve the discontinuity is called phase unwrapping. Methods [38, 39, 40] have been proposed to do the unwrapping. The basic idea is to convert the wrapped phase into the absolute phase [41]. The advantage of using phase shifting methods is that only three patterns are needed for the experiment, as represented by Eq. 2.2. The disadvantage is that phase unwrapping methods have to be introduced to resolve the discontinuity problem. Hence, phase shifting is not an ideal method to be used alone. Many methods are presented by combining the phase shifting methods with the Gray code projection methods [42, 43, 44].

2.2 3D reconstruction of transparent and specular objects

3D reconstruction of transparent and specular objects is difficult to achieve because of the active optical interaction of the objects with light, making the task very challenging to follow different paths of the light traveling through or reflecting from the objects. Obviously, a simple minded method, which is still used, is to cover the surface of the objects using an opaque powder. However, if the objects are delicate or precious, then it is not appropriate to obtain their 3D structures at the expense of damaging their surfaces. Over the years, researchers have come up with different ideas and methods to tackle the issues of transparency and specularity. Since it is impossible to cover all of these methods, only closely related and representative methods are reviewed in this section. In addition, the fundamental principles of 3D reconstruction of transparent and specular objects are conveyed.

For the 3D reconstruction of transparent objects, the existing methods can be roughly categorized into two groups. One group is the refraction-based methods, which mainly utilize the information of physical geometry and normally need more

than one viewpoint. The other group is the reflection-based methods, which generally apply to a larger range of applicable objects, because they do not have specific requirements for the objects' interior structures or properties. Here, the representative refraction-based methods are reviewed first.

2.2.1 Refraction-based methods

For transparent objects, most portion of the light gets refracted and transmitted through the objects. Only a small portion gets reflected. Comparing with reflection, refraction is much stronger. In addition, since the reflected light normally gets interfered by the reflection from the background, it is not easy to isolate the reflected light from the surface. Hence, many researchers measure the shape of transparent objects based on refraction only.

One of the main applications for refraction-based methods is in fluid surface reconstruction [45, 16]. The reason is because for transparent objects, the surface normally is specular, which means that there is one dominant reflecting angle. When the surface of the fluid varies, the dominant angle changes as well. Using the reflection-based methods, the viewing angle of the camera need to be adjusted accordingly, because the specular reflection is view dependent. Hence, for dynamic fluid surface reconstruction, the refraction-based methods [45] are more appropriate than the reflection-based ones [46].

Morris and Kutulakos [45] present a method using only two viewpoints to reconstruct a time-varying transparent fluid surface with no prior information about the refractive index of the fluid. A checkerboard pattern is located at the bottom of the tank. Using an optical flow estimation, a pixel-to-pattern mapping function can be acquired, under the assumption that the surface is composed of a homogenous transparent medium, and that the light gets refracted once only [45]. Using Snell's law, the 3D position and normal of the surface point at a certain time can be estimated. Then, the reconstruction error of all the points during the whole time range is computed and minimized. Through this optimization procedure, the refractive index value, as well as the 3D positions and normals of the points can all be obtained. Since the optical flow estimation is conducted, this method will fail when the change in the surface of the fluid is too drastic that the light does not arrive at the bottom (pattern) of the tank, instead, it may go to the side wall of the

tank. The major disadvantage of this method is that it only works when the light refracts once only, while in practice, the light has a much more complex interaction with the fluid and with the tank. Because of these strict requirements, their method has a very limited range of applications.

Kutulakos and Steger [47] present a theory to reconstruct the 3D shape of refractive and specular objects, which is called "Light-Path Triangulation". They unify the analysis of refractive and specular scenes, and enumerate all of the tractable light-path triangulation problems and show which ones are solvable [47]. For example, they point out that for light that undergoes two refractions or reflections, three viewpoints are enough for the reconstruction [47]. The results of their method rely on a pixel-to-pattern mapping function obtained by using an environment matting method. Their method also assumes the objects to have homogenous medium (opaque or transparent) and smooth surface (no surface scattering), which limits its applications. In addition, they have to move the cameras around the object, so as to get good viewing angles, which also complicates the image acquisition procedure.

Ding et al. [16] present a method which has a similar setup as [45], except that their new method requires more viewpoints. Since methods like [45] are known for their sensitivity to fast fluid motions, Ding et al. propose a new method to track the refracted feature points over time and across cameras, and obtain a spatial-temporal correspondence map [16]. Their method can efficiently acquire results with high resolution, and can track lost feature points by comparing the images from different cameras. However, since they need a 3×3 camera array, the calibration is more complex and error prone. In addition, adjusting all the cameras to focus simultaneously is very difficult and troublesome.

Aside from the 3D reconstruction for fluid, researchers have proposed many methods for reconstructing solid transparent objects. Trifonov et al. [11] introduce a visible light tomographic reconstruction method. The object is suspended in a fluid in a glass cylinder. The index of the solution is varied to match that of the object. Once the two indices are matched, the refraction at the interface between the object and the solution is minimized, and a tomographic technique is applied to perform 3D reconstruction using visible light, instead of X-ray. This method performs well for objects with an inner structure, such as a small hole in the center. However, it has several shortcomings. First, the solution used is potassium thiocyanate (potassium

salt) in water, which is highly poisonous. Not only is the solution dangerous for researchers to use, but also is the disposal very difficult after the experiment. The solution may also be corrosive or chemically active to the surface, which will damage the object permanently. Second, since they need to match the index of the solution with that of the object, it is not easy to find a proper solute that not only can make transparent solution but also can match the desired refractive index. Third, since they need to suspend the transparent object into a cylinder, the volume of the object is limited by the physical size of the cylinder. When the object is very big, this method will probably fail, because it is not easy to find a proper container to hold a big object and it will also need a large amount of toxic solution. In summary, although this method can obtain some good results, it is not practical to be widely used.

Wetzstein et al. [48] introduce a light field distortion method, which uses only one image to do reconstruction. The key is to develop a set of 4D spatio-angular light distribution patterns. Similar to some of the other refraction-based methods [45], this method assumes that the light refracts only once. They also assume that the refractive index of the object is known *a priori*, and the attenuation, the scattering, as well as the wavelength-dependency of refraction are neglected [48]. In addition, they can reconstruct only thin transparent surfaces, such as thin solid objects or the surface of fluids. Hence, this method also has a small range of applications.

Shan et al. [49] propose a method to obtain the refractive height fields by taking images with single or multiple planar backgrounds, with only one viewpoint. Their method, though can obtain accurate results, is computationally expensive.

To sum up, refraction-based methods tend to make assumptions such as the light is "refracted only once" [48], "the refraction index is known *a priori*" [11], and "the medium is homogenous" [45], to simplify their algorithms. Normally, these methods can only be applied to objects without or with simple interior structures and have simple interactions with light. Since these methods are fundamentally based on Snell's law, they either need to know the refractive index, or they need to develop additional methods to obtain it during the experiments. According to Kutulakos and Steger [47], for some scenarios, it is impossible to do refraction-based reconstruction, no matter how many cameras are used. However, for some scenarios, such as the dynamic fluid, when the reflection is too dim to capture, the

refraction-based methods are the only available methods.

2.2.2 Reflection-based methods

For the 3D reconstruction of solid transparent objects, comparing with the refraction-based methods, the reflection-based ones tend to apply to a larger range of objects and have less limitations and assumptions.

Matusik et al. [15] propose a novel method to acquire and render transparent and translucent 3D objects from arbitrary viewpoints under a novel illumination. They first acquire an α matte for each of the viewpoints, and then create the opacity hull accordingly, in order to accelerate the analyzing procedure of the following steps. Second, they use the method from [50], and obtain an environment matte. Third, they use the α and the environment matte to get a surface reflectance field, so as to composite any novel image from arbitrary angle. While their method can tackle both the refractive and the reflective effects, their setup is very complex. In particular, their method requires a plasma monitor, four light sources, six cameras, and two turntables, which not only makes their devices difficult to be mounted and manipulated, but also complicates the procedures of camera and system calibrations.

The scatter-trace photography method [17] was introduced by Morris and Kutulakos in 2007. The method can reconstruct the surface of transparent objects with complex interior structures. The goal is to get the depth and the normal for each point on the surface. A concept called "Scatter Trace of pixel q" is introduced and denoted as $T_q(L)$, which means that the incoming light at q when the light source is at point L [17]. Morris and Kutulakos display a 2-pixel-wide vertical stripe scanning through the monitor, and capture images of the object simultaneously. Since the monitor alone cannot determine the direction of the emitted light, the monitor must be moved along its screen normal to $3 \sim 6$ positions. The intensity of each point on the captured images is mapped to the corresponding light source positions. In this way, the highest intensities on the light source area indicate the direct component of the incident light, i.e. the light that is reflected by the surface of the object.

After acquiring the images and mapping the intensities to the light source area, the scatter-trace analysis for each point on the surface of the object is conducted. They first assign a hypothetical depth to the point on the object, and do the rectification, which is a linear projective warp that maps the point on the object

to an infinity point along the x-axis. In this way, the real scatter-traces that converge to the surface will be parallel to each other and to the x-axis, whereas the wrong ones will not be parallel to the x-axis. Based on this idea, their second step is to estimate the direct component and eliminate the wrong scatter-traces using a procedure called "running minimum" to remove indirect light that happens to be on that trace. The justification is that for a proper trace, the intensity must decrease monotonically in a radial direction away from the convergence point [17], which is a point on the surface of the object. After the "running minimum", if the point is opaque and receives few inter-reflection, they can see this point as having only direct component and only a single view is needed to do the reconstruction. In order to find the correct depth from all the hypothetical depths, they assume there is additive Gaussian noise and use the point-wise consistency to enhance the direct component and accumulate the enhancements across the 2D area of light sources. The maximization of the accumulation corresponds to the real estimated depth for this point. For a transparent point, since the indirect component cannot be neglected, two or more views are required to do the reconstruction. They modify the single-view method and evaluate the mutual consistency of scatter traces at corresponding pixels from different views [1]. For their implementation, instead of scanning a single stripe across the monitor, they use the patterned illumination multiplexing method [51] as the light source to reduce the acquisition time.

Morris and Kutulakos's method has a few shortcomings. For example, in their paper [17] and Morris's thesis [1], they fail to clearly state the method of rectification. In the thesis [1], Morris mentions that the rectification corresponds to the standard epipolar image rectification method, which is quite misleading, because it turns out to be not even close to epipolar rectification. After a few correspondences with Morris, it is clear that the rectification is actually a quadrilateral warping method. In addition, it is quite difficult to re-implement their experiments, because there are many requirements to meet. For example, the whole setup has to be inside a totally dark environment because the scatter-traces are very sensitive to environment light. That is because the reflection of a transparent object is quite dim with light sources emitted from an LCD monitor, and can be easily interfered by other light. Even the back-light from the LCD screen can affect the results. Second, the calibration is very difficult and yet very critical to the results. Inaccurate calibration result

will directly lead to bad experimental results. The camera has to be calibrated intrinsically and extrinsically. Morris et al. use a two stage calibration phase. They calibrate the camera to a calibration target first and then the illumination/monitor to that same target. Third, since the monitor has to be moved during experiments, high chances are that other part of the setup will be accidently moved too, which increases the possibility of errors. Fourth, since the monitor is moved $3 \sim 6$ times, the overlapped area that the light sources can cover all the times is very small. Hence, the reconstructed area of the object is very small, unless a turntable is used. Last but not least, this method is computationally time-consuming, because for each point on the surface of the object, they need to search the estimated depths in a large range in order to do maximization. However, since it is difficult to "guess" the range of depths, manual measurement is needed to estimate the distance between the camera and the surface of the object. Otherwise, it would be hugely time-consuming to do the arbitrary depth estimation. In summary, although Morris et al.'s method appears to be straightforward, the implementation is actually quite complex and difficult and probably irreproducible.

Meriaudeau et al. [9] propose a method of using emitted structured infrared patterns to reconstruct the nonopaque objects. They basically use a laser beam to heat up the object and use an IR camera to capture the emitted infrared patterns. Based on the temperature gradient at the laser intersection, they can locate the laser pattern corresponding to image pixels. The advantage is that they identify a new avenue to do 3D reconstruction using non-visible light spectrum emitted by the objects. The disadvantages are as follows. First, the mechanical setup is quite complex because they need a laser controller, a laser generator, a beam expander, and a cylindrical lens to create the desired laser plane, and the camera must be a special thermographic one. Second, since they need to project the laser plane and capture the scene with the IR camera alternatively, the experiment is quite time-consuming. Third, although this method does not contact the object physically, it does heat up the object constantly, which will potentially melt and damage the object. Fourth, since the IR camera has a low resolution, the results are less accurate than a typical structured light method. Last but not least, infrared patterns are quite sensitive to heat noise, such as the heat from the environment. Hence, special precaution to control environment heating must be used.

In addition to the methods reviewed above, there is another group of reflection-based methods, which is the polarization-based methods [12, 13, 52, 53]. These methods make use of the polarization phenomenon and by measuring the degree of polarization, they can obtain the reflection angle and the surface normal of the object. The challenge is to overcome the ambiguity problem of the two surface normals. Miyazaki et al. [12, 13] propose a solution for this problem, which is to tint the object for a small angle and capture additional images from this new view. By comparing the images from two views, the orientation problem can be resolved. This method uses only the sign, instead of the value of the rotation angle. Hence, they do not need to know the actual value of rotation, nor need to do calibration. For simple objects with a smooth surface, this method can acquire good results, but for objects with unknown refractive index, or objects that self-occlude, this method may fail.

Generally for polarization-based methods, since unpolarized light sources are needed to illuminate the object from all directions, the object normally needs to be put inside a diffuser, such as a plastic sphere. In this way, the light from the light bulbs outside the diffuser can be modified to be the desired form. Inside the diffuser, there are some inevitable light effects such as inter-reflection. Since the shape of the object is unknown, it is quite complicated to deal with these inter-reflections [12]. Hence, some researchers [12] simply assume that the light caused by inter-reflection is an uniform unpolarized light, and subtract its intensity before further analysis. Nevertheless, even with this strategy, the errors of the results are still mainly caused by inter-reflection, which indicates that this strategy still has problems.

Some of the methods mentioned above can also be used to reconstruct the specular objects, since many transparent objects are also specular. It is noteworthy that most of these methods are reflection-based, because refraction-based methods normally need to put pattern(s)/monitor(s) and camera(s) at different sides of the objects, and light normally transmits through the object into the camera(s). Furthermore, the refraction-based ones need more information about the objects, such as the refractive index. In addition to the reflection-based methods mentioned above, there are several methods which are designed specifically for specular objects.

2.2.3 3D reconstruction of specular objects

Nehab et al. [54] introduce a dense 3D reconstruction method for specular objects based on the surface normal/depth consistency. The setup consists of a monitor displaying scanning stripes as the light sources, with two cameras capturing image pairs at the top of the monitor. The limitation of the setup is that since the consistency property is used, the object need to be located pretty close to the light sources. Hence, the angle covered by the monitor is very small. This method can only performs well when the surface is nearly flat, otherwise, "gaps" can be observed in the reconstruction results. In addition, even after the inter-reflections are neglected, the consistency constraint can still lead to ambiguities. In their paper, they present a theoretical analysis of the ambiguities.

Different from Nehab et al. [54], who display scanning linear light sources on the monitor, Francken et al. [55, 56, 57] propose a few coding methods so as to reduce the acquisition time and to improve the accuracy of the results. Since a specularity can be caused by only one light direction, once the point on the monitor is located corresponding to the specularity, the local normal can be easily acquired. Francken et al. [55, 56, 57] chose the Gray code patterns to display on the monitor, since the Gray code patterns are very robust to noise and to errors. Apart from the Gray code patterns, they also use the gradient illumination patterns. However, this pattern is more sensitive to noise than to the black/white binary patterns.

2.3 Environment matting

As discussed above, the main challenge for 3D reconstruction is to find the correct correspondences for triangulation. While existing methods for 3D reconstruction of transparent and specular objects have certain limitations to tackle this problem, there is another research area called "environment matting", whose goal is to determine the interaction between a transparent object and its environment, has made many exciting progress in finding the correct correspondences. The proposed method of this thesis is one of the first ones to adapt a method in environment matting to 3D reconstruction. Before going into details about how the proposed method modifies an environment matting method and uses it as the first step of the reconstruction process, relevant background of environment matting and related methods

are presented and reviewed below.

2.3.1 Background: matting

Environment matting is closely related to image matting. Matting and compositing are originally invented for film production [4]. Sometimes when it is quite difficult to shoot the film in a scene, for example, a scene which is setup on the Moon, people would shoot the scene in front of a big blue or green screen. The blue or green screen is the "background." The people in front of it are the "foreground." Using a matting method, the foreground can be separated from the background and be composited with a new background.

Porter and Duff [58] are the first ones who introduce the idea of alpha channel and matting. They separate the image into different parts which can be independently rendered. Each part has an associated matte, so that they can render the scene separately and accumulate them into an image. In particular,

$$C(x,y) = \alpha(x,y)F(x,y) + (1-\alpha(x,y))B(x,y) \tag{2.4}$$

where C(x,y) is a pixel on the image, α the alpha value. F(x,y) the foreground pixel, and B(x,y) the background pixel. Ideally an image can be strictly separated into the foreground region and the background region, i.e. one pixel can be either in the background or in the foreground. However, in most cases, the image has three parts instead of two, and that image is called a *trimap*. Fig. 2.4 shows a trimap. The blue part (labelled as "1") is the definite background, and the corresponding alpha value is zero. The yellow part (labelled as "2") is the definite foreground, and the alpha equals to one for each pixel in this area. The gray area (labelled as "3") is the uncertain area, where the pixels can be both the foreground and the background, and the alpha value is between 0 and 1. From the viewpoint of the trimap, "matting" is actually a generalization of image segmentation. When the object has a sharp edge and is totally opaque, α is either 0 or 1, and matting becomes image segmentation. Matting is widely used when the foreground object does not have a hard edge, such as fur or hair.

Matting is an ill-posed problem. Given that an image has RGB channels, there are seven unknowns in Eq. 2.4. The foreground and the background pixels each have RGB three unknowns, and the alpha value is also unknown. There are only

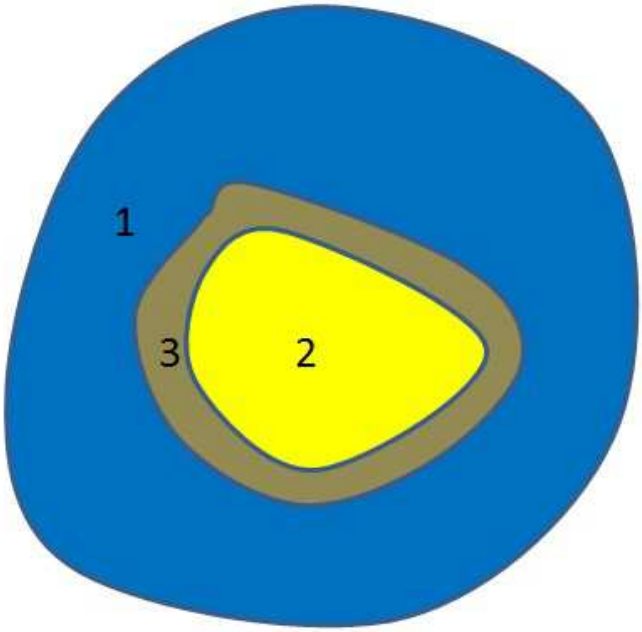


Figure 2.4: Trimap.

three equations (for the three RGB channels) to solve for seven unknowns. Since the unknowns are out-numbered, it is quite difficult to solve them and most existing methods [59, 60, 61, 62] use optimization or user-interaction to solve this problem.

2.3.2 Environment matting and compositing

Matting can only be applied when the foreground object does not interact with the environment, which means that the captured image is a composition of the weighted foreground and the weighted background. When the foreground object is transparent, translucent, shiny, glossy, or seen at a grazing angle, the environment light can have an active interaction with the object, causing more optical effects to be analyzed, such as reflection, refraction, attenuation, scattering, inner reflection, inner refraction, and inter-reflection.

Since using only the theory of "matting" is not enough, Zongker et al. [63] first introduced the concept of "environment matting" in 1999 to represent these optical effects mentioned above. They propose the "environment matting equation" as

$$C = F + (1 - \alpha)B + \Phi, \tag{2.5}$$

where α only represents the coverage of a pixel by the foreground, while in the original matting equation, Eq. 2.4, α represents the coverage, as well as the opacity

of the foreground pixel [63]. Φ represents the contribution of any light from the environment that reflects from or refracts through the foreground.

Since Φ is quite complex to represent and to calculate using the captured images, Zongker et al. make a few assumptions to simplify and to approximate Φ, such as "the only light reaching the foreground object is the light coming from distant parts of the scene" [63]. Based on these assumptions, a simplified model of light transport is presented as

$$C = F + (1 - \alpha)B + \sum_{i=1}^{m} R_i \mathcal{M}(T_i, A_i) \text{ [63]}, \tag{2.6}$$

where R_i is the ith reflectance coefficient and T_i is the corresponding texture map. They assume that there are a set of m different texture maps, denoting contributions of the light from m different parts of the environment, such as different angles. $\mathcal{M}(T_i, A_i)$ is a texture-mapping operator that returns the average value of an axis-aligned region A_i of the texture T_i.

The goal of environment matting is to solve the equation above. Since there are RGB three color channels, there are three unknown foreground colors for F, three unknown corresponding reflectance coefficients for R, four unknown area extents for each A_i, and one unknown pixel coverage value for α, at each pixel. Using coded patterns displayed from the background and to the sides of the object, and with a non-linear optimization procedure, these unknowns can be solved.

This method, though is pioneering, has a couple of shortcomings. For example, the number of images required to capture an environment matte is large, which is space and time consuming. In addition, according to their setup, the side-drops are actually only put to the left and right of the object, whereas in reality, the environment light can come from everywhere around the object. Hence, their setup does not cover the whole range of the environment. Last but not least, this method can not handle abrupt changes in reflected and refracted ray directions.

In 2000, Chuang et al. [50] proposed a few extensions to their previous work [63]. The extensions are in two directions. One is aimed at a higher accuracy, and the other attempts to achieve a higher efficiency. In order to obtain the results with higher accuracy, they develop a new set of patterns. Instead of using the Gray code patterns in structured light, they use a single scan-line to scan through the monitor. The scan-line comes in four directions, vertical, horizontal, and two diagonal di-

rections. In addition to their improvements, instead of using simplified rectangular regions as weighted function, they use the sum of Gaussians as the new weighted function, which can allow for a simulation of dispersion and also a simulation of convergence. The results of this method are much better than their old ones. However, since more patterns are used, it is more time consuming than the previous one [63].

The other direction of their extensions is higher efficiency. They simplify their object into moving, deforming colorless ones with specularly reflective and refractive properties and utilize only a single background image consisting of a color ramp for real-time capturing [50]. Although their method is more efficient, because of the assumptions of the object, it is limited to a small group of objects. However, a road-map to simplify the environment matting is proposed, which has inspired other researchers [3, 64, 65, 66].

2.3.3 Efficiency-based methods

Ever since Zongker et al. introduced the idea of "environment matting" and proposed an inspiring method to obtain the environment matte in their paper, researchers have come up with many new methods and ideas for environment matting. Similar to the work of Chuang et al. [50], these new methods can be roughly categorized into two parts, based on their main goals. One part is efficiency-based methods, and the other part is accuracy-based methods.

There is always a trade-off between efficiency and accuracy. Although the ultimate goal is to achieve both. In practice, researchers have to emphasis one over the other when developing their own methods. Some researchers [3, 65] make a few assumptions to simplify the environment matting model, so that their methods can be efficient, yet can only be applied to a limited range of objects or scenarios. On the other hand, some researchers [4, 67] develop a novel set of patterns to display on the monitor, so that they can accommodate as many optical effects as possible into account and obtain more accurate results, yet may take hours to do the experiments. Here, some state-of-the-art efficiency-based methods are reviewed.

Wexler et al. propose a method to use only images with a moving background to obtain the environment matte without calibration. In the paper of Zongker et al. [63], one shortcoming of the method is that the acquisition process is not accurately calibrated, which leads to inaccurate results. As well, Wexler et al. [3]

point out that for some scenarios, for example, when the transparent object cannot be moved to or too big to be placed inside a setup, it is quite difficult to do the calibration. By avoiding the calibration, the experiment can be accelerated as well. The main idea is shown in Fig. 2.5, where all the pixels of the background image have contributions to each pixel on the composited image, though a certain number of these contributions may equal to zero. For each pixel on the composited image, a contribution map, which is called the "receptive field" in Fig. 2.5, is created to record the corresponding contribution of every pixel from the background image. The sum of the weighted background is the intensity of the pixel on the composited image. In case of no prior information of the background, a technique is presented to generate "a clean background" by comparing a series of images with the overlapped background. The accuracy of this method increases as more images are available.

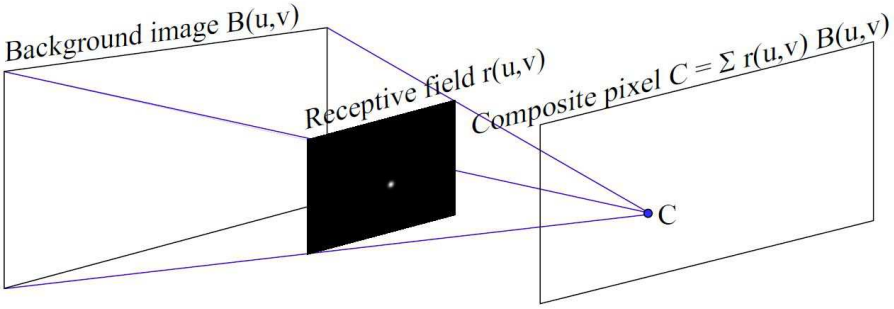


Figure 2.5: Formation of a single output pixel [3].

The disadvantages of this method are as follows. According to the main idea stated above, only the contributions from the background image are considered, whereas the method in [63], the environment light can be from anywhere, which means that Wexler et al. simplify the "environment" to be the background only. Hence, their results do not have reflection from the environment and only experiments with optically simple objects are shown, such as a thin magnifying glass. In addition, in the process of getting the foreground elements, they assume that the foreground pixels have lower variance than the background ones, which may fail when the foreground moves faster than the background. Moreover, their method may fail when the background is non-planar.

Choudhury et al. [68] introduce a method to efficiently acquire the environment matte using a holistic color cube as the environment. Different from the method proposed by Wexler et al. [3], this method takes light coming from every angle into

account. In order to find which point on the cube corresponds to a certain pixel in the final image, instead of lighting one point at a time, they use color as a cue to locate it. There are mainly two disadvantages of this method. First, it can only deal with purely reflective/refractive objects, and may fail when the object has a complex interaction with the environment. Second, since the foreground object is in a cube, the size of the applicable objects is limited.

Inspired by the work of Chuang et al. [50], Duan et al. [64] propose a follow-up method to remove the potential errors in the compositing results, which improves the result of [50] in accuracy. Based on their theory [64], Chuang et al.'s method can only remove high frequency noise, but not errors, which can be seen as the low frequency noise. Duan et al. use a concept called "light motion vector," which is a vector starting from a point on the foreground to the corresponding background. These light motion vectors form collectively a light motion field. The efficiency-based method proposed by Chuang et al. [50] is adopted to obtain the light motion vectors. An energy minimization method is introduced to remove the noise and errors at the same time. This method is very efficient. However, since the method from [50] is used, the same assumption of the object has to be made, which means that the object has to be colorless and transparent. Although this method is fast and has a higher accuracy than the previous method [50], it is still limited to a small group of transparent objects with the colorless property.

In 2011, the same group of people, Duan et al., came up with a new method [65] which adopt the compressive sensing method in signal processing, and obtain good results. The main assumption of compressive sensing is that the input signal is sparse. With the help of compressive sensing and group clustering, the environment matting model is simplified and hierarchically solved. This method can obtain accurate results very efficiently. However, since the input signal is assumed to be sparse, this method can only acquire the environment matte of simple transparent objects. Otherwise, when there are more optical effects, such as scattering, caustics, inter-reflection, and attenuation, and these effects may come from many points from the environment, the input signal may not be sparse anymore. In addition, in order to use clustering to group the foreground pixels into a few classes, local smoothness is assumed. Hence, this method cannot deal with objects that have abrupt changes or complex structures.

By far the methods reviewed in efficiency-based category have at least one thing in common, which is the applicable objects are simple and have little complex interactions with the environment light. This is quite easy to understand. Since the goal of these methods is to achieve a higher efficiency, assumptions or limitations have to be imposed so that the environment matting model can be simplified and the capturing procedure can be accelerated.

However, not all of the efficiency-based methods are restricted to be applied to simple transparent objects only. Yeung et al. [66] propose a novel method that can accomplish matting and compositing in almost real-time, and yet the applicable objects can be optically active and complex. In particular, their method is derived from the traditional matting for opaque objects, instead of deriving from the environment matting methods proposed by Zongker et al. [63]. However, the foreground object they use is also transparent and refractive. Their method can produce realistic effects that previously only 3D modelling and environment matting methods could achieve [66]. Instead of using the environment matting model, they propose a novel model called "Attenuation-Refraction Matte," or ARM. Taking the advantage of our visual tolerance, and with a few reasonable assumptions, they come up with a discrete form for the image formation equation as

$$C_M(x) = \alpha(x)S(x) + (1 - \alpha(x))\beta(x)B(G(x)), \tag{2.7}$$

where x is a pixel located within object M, α the relative contribution of the specularity S, β a 3-channel color transmission factor, B the appearance of the background without an object, and G the warping function [66]. The main task of this method is to solve for Eq. 2.7 and use the obtained M, α, β, and G to composite new images with arbitrary backgrounds. Since using only one image to solve for Eq. 2.7 is an ill-posed problem, they make further assumptions to simplify the acquisition of α, β, and G, and use user-interaction to extract other unknowns.

The way to assess their results is quite unique. They ask a *Photoshop*® expert to do matting and compositing using *Photoshop*® and ask a group of random people to compare the results of the ARM approach with those of the *Photoshop*®. According to their survey, the ARM approach generally produces more preferable results. However, this is the only assessment available of their results, and there is no quantitative assessment.

The advantages of this method are that many transparent and refractive objects can be used to extract their environment matte, even with complex structures. The matting and compositing procedure is nearly in real-time, and visually plausible results are acquired. The disadvantages are that user-interactions are required during the whole process. The results are only visually plausible, and no mathematical or physical information about the surface of the object is obtained, nor the mapping from the background to the composited image is acquired. In addition, since they use user-interaction during the analysis, their method cannot be applied to environment matting from videos.

In summary, although the efficiency-based methods are quite efficient and sometimes even in real-time, they tend to make assumptions to simplify the environment matting model and to accelerate the capturing procedure. They normally have limitations for applicable objects and environment scenarios. Sometimes they can only get visually plausible results, without getting real physical information about the structure or optical properties of the object.

2.3.4 Accuracy-based methods

According to the previous reviews, efficiency-based methods normally can be only applicable to simple transparent objects which have simple or unique interaction with the environment light. These methods normally will fail when the foreground objects have a complex interior structure or surface. Over the years, researchers have studied and proposed many accuracy-based environment matting methods that can be applicable to a more general group of transparent objects.

Among these accuracy-based methods, Zhu and Yang [4] introduce an elegant method called the frequency-based environment matting method. This method is inspired by the fact that a time domain signal has an unique decomposition in the frequency domain. Optically complex objects incline to converge or disperse light paths when put in front of the background. This phenomenon is quite difficult to simulate because there are infinite ways to decompose a pixel [4], i.e. it is very difficult to determine where the light paths start from based on the captured images. Previous methods [50, 63] use non-linear optimization to solve this problem, but the solution is only an approximation, which is not accurate. Since in the time domain, the possibilities of decomposition are countless. In the method proposed by Zhu and

Yang [4], the signal is transformed from the time domain to the frequency domain and a unique decomposition is obtained. The advantage of this method is that it avoids the optimization step, which cuts down the processing time. In addition, since the mapping result shows the actual decomposition of each pixel, the compositing result with an arbitrary background is physically correct, rather than only visually plausible.

Since the proposed 3D reconstruction method in this thesis is derived from [4], this method [4] is reviewed in detail here. The setup is quite typical, with the camera in front of the transparent object and a CRT monitor in the background. The frequency of a signal is only determined by the source that created it, and is not affected by the medium when moving through. As shown in Fig. 2.6, the two pixels on the pattern with different frequencies transmit and get converged into the same point in the captured image. Although the converged point has a mixture of signals, it actually can be decomposed into two unique signals in the frequency domain. These two signals, respectively, correspond to the signals from the two points on the pattern.

Since the goal is to use frequency to uniquely locate the decomposed components, each position on the patterns should have a unique frequency. The pattern has a 320×320 resolution. To accelerate the image acquiring procedure, two groups of patterns, row-based and column-based, are designed. Because the signals with low frequencies are easier to get interfered and to be mistaken as noise, the upper-left point is not set to have a 1Hz frequency, instead, it is set to be 11Hz. Hence, the range of frequencies on the patterns is between 11Hz and 330Hz. Based on the Nyquist Sampling Theorem, at least 660 patterns are needed. In their experiments [4], 675 row-based patterns and 675 column-based patterns are used.

The environment matting model can be represented by

$$C(x,y) = F + R \sum_{s=1,t=1}^{s=S,t=T} W(s,t)B(s,t) \qquad where \sum_{s=1,t=1}^{s=S,t=T} W(s,t) = 1, \tag{2.8}$$

where $C(x,y)$ is the pixel on the output image, F the color of the foreground object, $B(s,t)$ the pixel (s,t) on the background image, which has $S \times T$ pixels in total. Similar to [3], W is the normalized weight map showing the contribution of every background pixel to a certain output image pixel. More specifically, W shows not

only which background pixels contribute to the output pixel, but also the proportion of their contributions. R is the reflectance coefficient, showing the attenuation of the light when interacting with the foreground object. The goal is to solve for F, R, and most importantly, W.

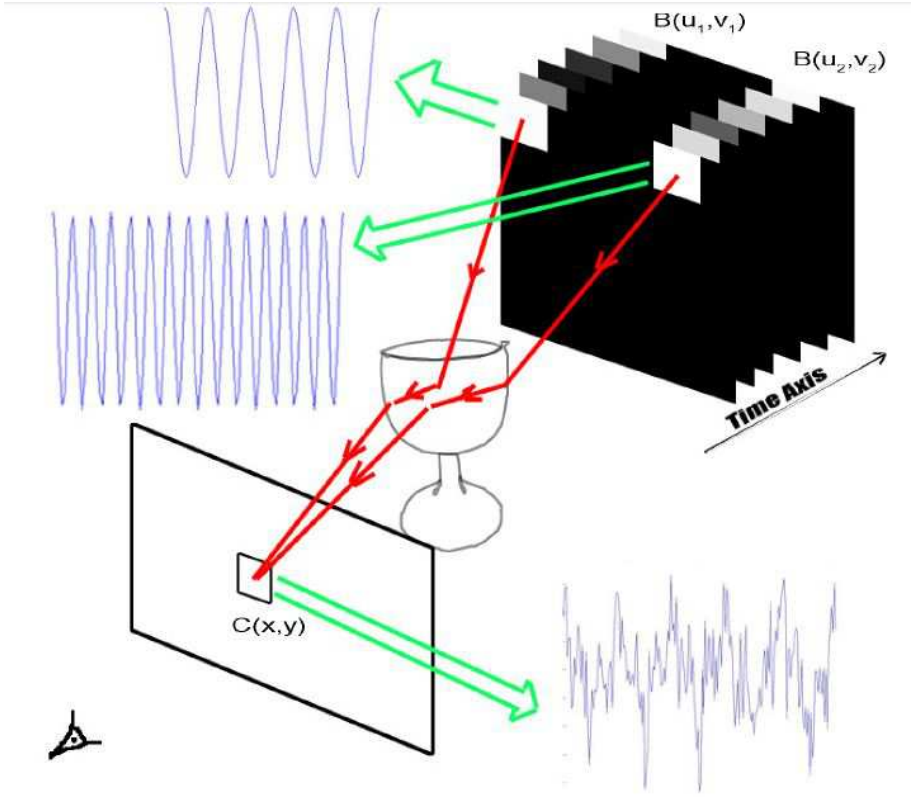


Figure 2.6: An illustration of the frequency-based environment matting method [4].

To acquire the environment matte, first the same method proposed in [63] is used to get an α matte of the object. Second, the solid black pattern is displayed on the monitor so as to get the information of the foreground color. Third, the foreground color is used to obtain the reflectance coefficient. Fourth, the patterns are sequentially emitted by the monitor and the camera takes the images simultaneously. Since the monitor, the object and the camera do not move, each pixel on the pattern has a unique path that does not change with the patterns. When the set of captured images are "seen" along the time axis, a sequence of intensities for each pixel is obtained. The sequence is transformed to the frequency domain using the Discrete Fourier Transform. In the frequency domain, the signal is analyzed, and the dominant frequencies are acquired. Each of these frequencies corresponds to a unique position on the pattern. The coefficient of the frequency indicates the proportion of the contribution to the pixel on the final image. In this way, the

weight map W can be acquired. For the compositing step, since the values of the unknowns have all been acquired, it is easy to use Eq. 2.8 to compose a new output image.

Since the frequency analysis is robust to noise, the frequency-based environment matting method performs well even interfered by noise. The authors intentionally add noise to the captured images, and still get very good results. Even after intentionally eliminating a couple of captured images, their method still works well with the rest of images. Because of the good properties of the frequency domain analysis, the foreground object can be considered as located inside a black box, and there is no need to know the interior structure or the surface of the object. Hence, this method can be widely used for a large range of objects, even objects that have a complex interaction with the environment light. However, the disadvantage of this method is low in efficiency. Although this method can acquire accurate results, it is very time-consuming to capture all the images. Normally about 30 minutes are needed to do one experiment.

Peers et al. [67] also propose an accuracy-based method. Instead of using the sinusoidal Fourier Transform patterns, they use the Haar wavelet patterns. The advantage of the wavelet patterns is the local support in both the time domain and the frequency domain. Hence, there is no need to do the signal transformation, but the signal decompostion of the wavelet is needed. Since the wavelet can be decomposed iteratively, a hierarchical procedure is developed to save time. In addition, an errortree is introduced to indicate which wavelet to emit next. Only the most important one is chosen to emit. During the procedure, the contribution of the wavelet to the illumination of the scene is calculated, and the result is stored in an errortree. Then the next level of the wavelet is searched in the errortree and the chosen wavelet is emitted. This procedure is conducted continuously in a loop, and stops when the acceptable result is reached or when the number of the iterations is exceeded.

This method can also obtain very accurate results. However, user-interaction is needed to determine the stopping criterion, which is not reliable. In addition, in the feedback loop, they first calculate the contribution, and based on the contribution, they determine which wavelet pattern to emit next. Hence, it will take a long time to finish the experiment. Normally, the image capturing procedure and processing

procedure are separate. For example, in [4], though they have 1350 images to take, their image processing time is very short, normally 10 minutes. Hence, Zhu et al. need 2 hours in total to do the experiment. However, since Peers et al. do their image capturing and processing alternately, they need 12 hours in total for each scene and an average of 2.5GB to store all the photographs (after compression) [67]. In summary, their method gives good results but is very time and space inefficient.

Generally for the environment matting methods, the devices for displaying the patterns are LCD, LED monitors, or plasma panels. There are some shortcomings about these devices. First of all, since these devices are used as the light sources, the light emitted by the devices sometimes is quite dim. Especially when the object is far from the device or when the environment light is very strong, the quality of the captured image will be inevitably affected. Second, using these devices can only show the information of the corresponding points, without the directions of the emitted light. For environment matting, since the patterns are displayed at the background, and the camera captures the images in front of the object, most portion of the light that emitted by the monitors or the panels goes into the camera. Hence, the intensity of the captured image is not weak. However, when adapted to be used for 3D reconstruction of transparent and specular objects, not only the reflected light is only a small portion of the emitted light so that the intensity of the captured image is very weak, but also using monitors or panels that do not indicate the information of the emitted direction, which is much more difficult than using a structured light system. Morris et al. [17] provide an idea, which is to move the monitor to $3 \sim 6$ positions, and since the moving direction is vertical to the surface of the monitor, a 2D rectangular range of stripe light sources can be created, and the corresponding information from different positions can be used to obtain the direction from the point on the monitor to the point on the surface of the object. This method, however, is not practical, since moving a monitor during an experiment can introduce systematic errors, and makes the calibration very difficult, and sometimes even not doable. One solution to this problem is to use structured light as the light source. The advantage is that the intensity emitted is much stronger and the direction from the projector center to the pixel on the pattern can be easily acquired, which is quite important for triangulation.

Out of space concern, only two accuracy-based methods are reviewed. The main

stream of environment matting methods are the efficiency-based ones. However, when researchers want to utilize the environment matting methods into other research areas, a method that is widely applicable to various transparent objects is more favored, because researchers tend to get good results first, and then make it to be more efficient. That is the reason why the proposed method of this thesis modify the frequency-based environment matting method and use it as its first step.

2.4 Conclusions

Traditionally for the 3D reconstruction of opaque objects, using structured light is the prevailing method. Researchers have developed many coded patterns so as to reduce the acquisition time and space, as well as to increase the accuracy and resolution. However, for transparent objects, because of their interaction with light, the traditional structured light methods can not produce persuasive results.

3D reconstruction of transparent and specular objects can be roughly categorized as the refraction-based methods and the reflection-based methods. Comparing with the reflection-based ones, the refraction-based methods normally have multiple assumptions so as to simplify the reconstruction model to make it solvable. They also need to know the properties of the object, such as the refractive index. In addition, the objects they can reconstruct normally have a very simple or no interior structure, and have a very simple interaction with light. Hence, the refraction-based methods normally have more limitations than the reflection-based ones. However, for a transparent object, if a refraction-based method can be applied, the reconstruction result is quite desirable. Generally, the reflection-based methods have a larger range of applicable objects than the refraction-based ones.

Although transparent and specular objects have complex interactions with light, researchers in the environment matting area have come up with some good methods to tackle these problems. The frequency-based environment matting method can accurately find multiple correspondences for each pixel in the image, with their weight map denoting the percentages of the contributions. With the help of the environment matting methods, the traditional structured light methods can be used to do the 3D reconstruction of transparent and specular objects efficiently and accurately. Hence, the frequency-based environment matting method can be modified and used as the first step in the proposed 3D reconstruction method, so as to help

find the correct corresponding points on the projected patterns with the pixels in the captured images.

Chapter 3

Frequency-Based 3D Reconstruction of Transparent and Specular Objects

3.1 Introduction

3D reconstruction of transparent and specular objects has been an active topic for many years. It has been widely used in medical science [69, 70], industry [71, 72] and entertainment [73, 74]. Because of the active interaction of the objects with environment light, these objects are difficult to be reconstructed. Traditionally, methods for 3D reconstruction of opaque objects use structured light with coded patterns, but these methods may fail for transparent and specular objects because the projected patterns may get reflected by the background and interfere the real reflection from the surface of the object, making it difficult to find the correct correspondences between the pixels on the projected patterns and the pixels on the captured images. However, the frequency-based environment matting method [4] can be adapted to accurately find the correct correspondences.

The transparent and specular objects interact with light in a complex fashion. Since the object may have a complex interior or exterior structure, the light may get reflected or refracted multiple times before it comes into the camera. Using the frequency-based environment matting method, a many-to-one mapping matrix from the pixels on the projected patterns to the pixels on captured images can be obtained, which denotes the convergence of multiple light paths. The many-to-one mapping matrix not only contains the points on the surface, but also contains the points off the surface. The points on the surface are generated by the first-

order reflections, which are reflections at the closest surface to the camera. This is similar to the traditional 3D reconstruction of opaque objects, as shown in Fig. 3.1. The points that are off the surface are generated from triangulation shown in Fig. 3.2. This scenario happens when the emitted light from the projector gets refracted into the object and after multiple inner reflections and refractions, the light gets refracted out of the object and goes into the camera. Since using the frequency-based environment matting method can only obtain the corresponding points from the projector plane to the camera image plane, triangulation using the incoming direction and the outgoing direction may converge at a point that is not on the surface, as shown in Fig. 3.2. In practice, the points off the surface are very common because of the complex interaction of the object with light. Since the goal is to reconstruct the surface of the object, only the points acquired from first-order reflections are detected and preserved, whereas points off the surface are eliminated. This procedure is done using the labelling method proposed in this thesis.

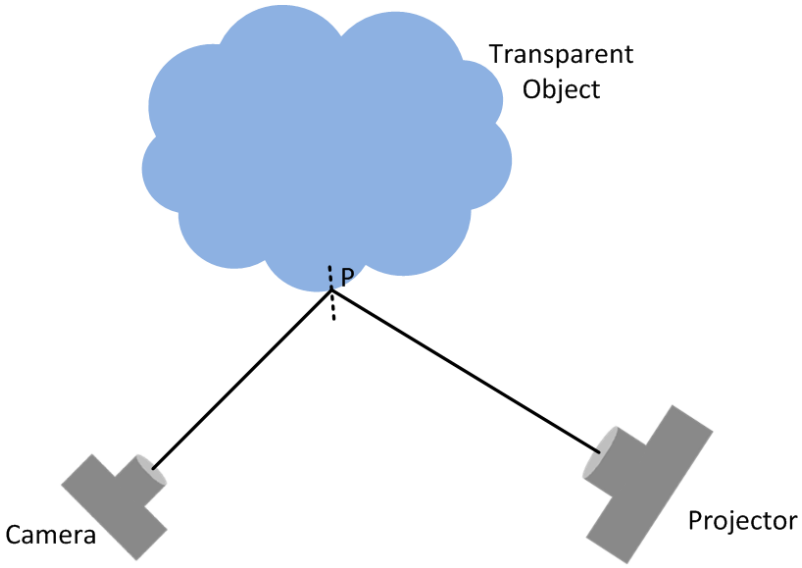


Figure 3.1: First-order reflection. Other optical effects are not illustrated here.

Once the converging points from first-order reflections are calculated using triangulation, the surface of the object can be assembled. However, since a projector normally has a lower resolution than a camera, the patterns projected onto the object tend to cover more than one pixel in the image. Hence, the postprocessing methods are required for refining the results.

In the next section, the main method is described step by step in details, and the related knowledge of this method is presented as well.

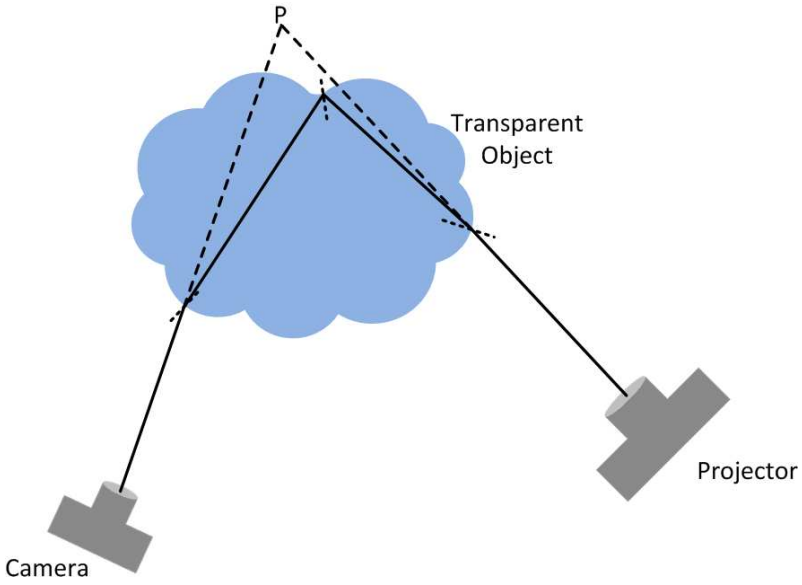


Figure 3.2: Points generated by triangulation may not exist. Other optical effects are not illustrated here.

3.2 Frequency-based 3D reconstruction of transparent and specular objects

3.2.1 Overview of the method

To clearly explain this method, a few things about the setup are needed to be explained first. Shown in Fig. 3.3, the experimental setup is similar to the traditional setup for 3D reconstruction of opaque objects. The projector and the camera are located on the same side of the object. The object is put before a black cloth, to minimize the interference from the reflected light from the background. Since the reconstruction is based on reflection, the relative positions of the projector, the camera and the object need to be adjusted, so that the camera can receive as much reflected light as possible.

3.2.2 Environment matting

As mentioned in Chapter 2, the first step of the proposed method is modified from Zhu and Yang's frequency-based environment matting method [4].

Frequency analysis

Same as [4], in the proposed method, the intensities of each pixel position in the captured images are "seen" along the time-axis as a sequence of signals. Since these

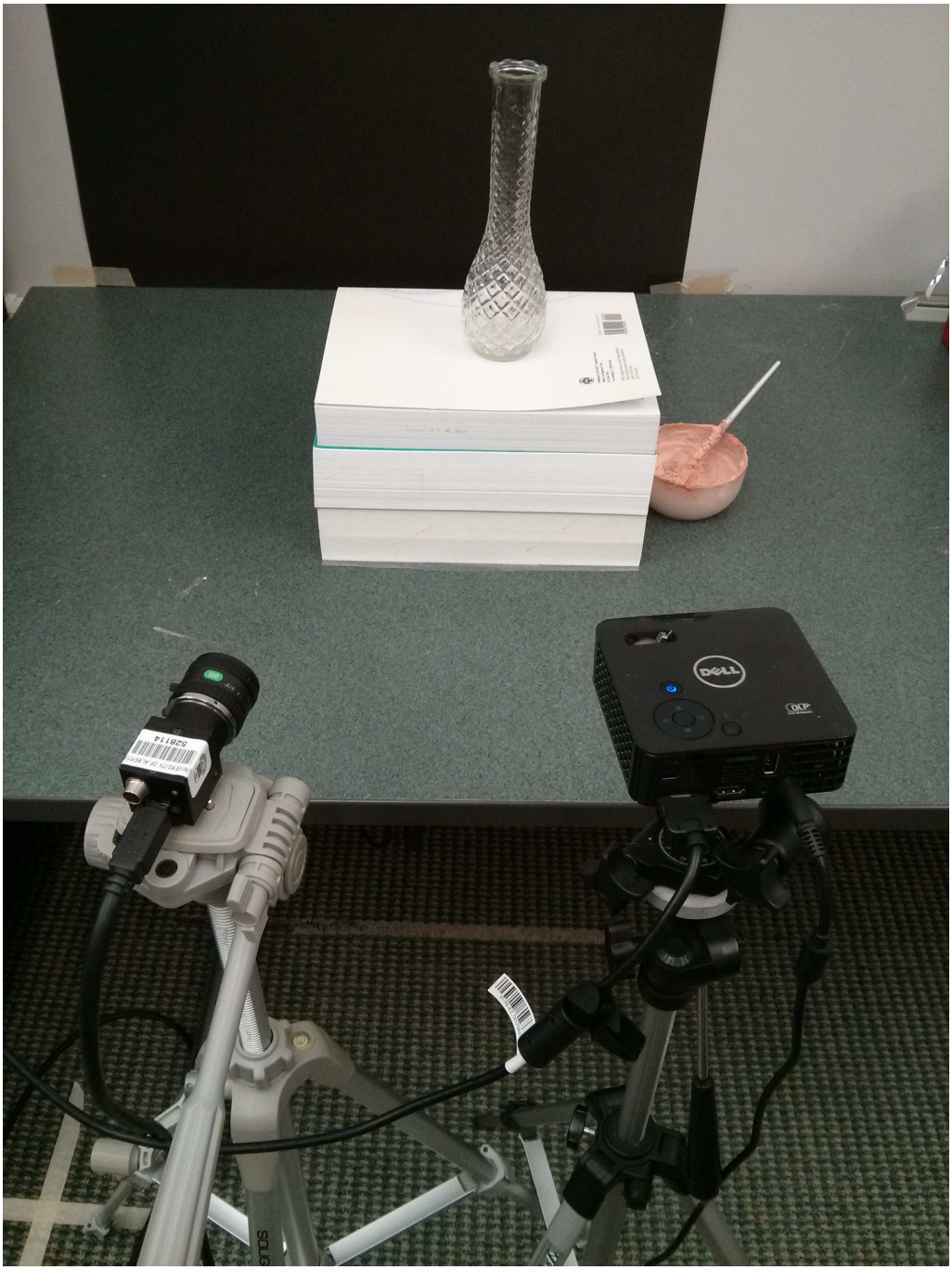


Figure 3.3: Experimental setup.

signals are a series of discrete numbers, the Discrete Fourier Transform, also known as DFT, is used to do frequency analysis. DFT can be represented by Eq. 3.1 [75]

$$X_k = \mathcal{F}\{x_k\} = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi kn/N} \qquad for \quad k = 0, 1, ..., N-1 \qquad (3.1)$$

where a sequence of N complex numbers $x_0, x_1, ..., x_{N-1}$, denoting a series of discrete signals in the time domain, is transformed into an N-periodic sequence of complex

numbers $X_0, X_1, ..., X_{N-1}$, denoting the signals in the frequency domain [75].

The key property of frequency, which is also the reason for using the frequency domain analysis, is that the frequency of a signal only relies on the source that creates it, and does not change when moving through the medium. Hence, when the projector casts a series of patterns onto the object, the pixel in each position of the pattern forms a "signal" and interacts with the object in a complex way. For example, it may directly get reflected on the surface, or it may refract into the interior and get multiple reflections, and then refracts out of the object. However, no matter how complex the interaction may be, the frequency of this signal remains unchanged. The signal is formed by a series of intensities, and as long as the intensities come in a periodic way, its frequency can be easily calculated using the Fast Fourier Transform (FFT) [76]. Fig. 3.4 is an example of using the FFT to obtain the frequency of a signal.

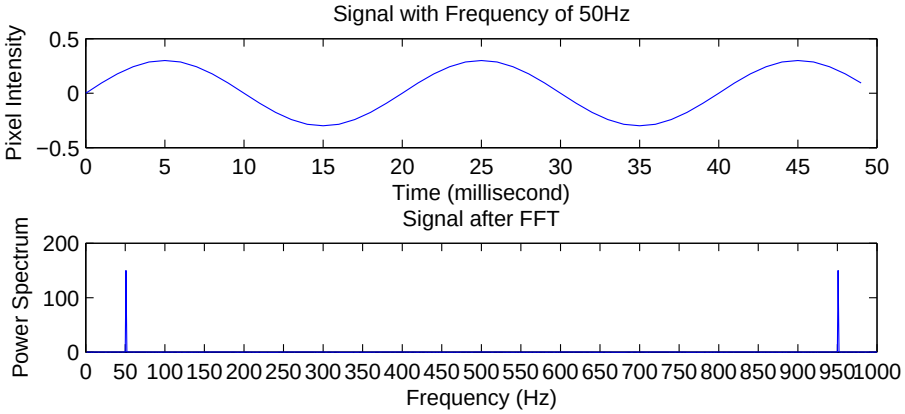


Figure 3.4: An illustration of using FFT to obtain frequency.

Fig. 3.4 shows a sinusoidal signal $f(t)$

$$f(t) = 0.3sin(2\pi \cdot 50t) \tag{3.2}$$

with a frequency of $50Hz$. It can be transformed from the time domain into the frequency domain and its frequency is prominently shown. However, this is not the only advantage of frequency analysis.

According to [4], there are three additional desirable properties of frequency analysis that can benefit the correspondence process:

1. Although a signal may be a sum of multiple signals with different frequencies,

transformed into the frequency domain, the frequencies of these components clearly show up.

2. Frequency analysis is robust to noise, which means that when a signal is interfered by noises, even though in the time domain, the signal is highly affected by the noise, when transformed into the frequency domain, the frequency of this signal is unchanged.

3. When the signal is scaled, its frequency domain form preserves the same components of frequencies, and their power spectrum are proportional to the square of the scaling factor [4].

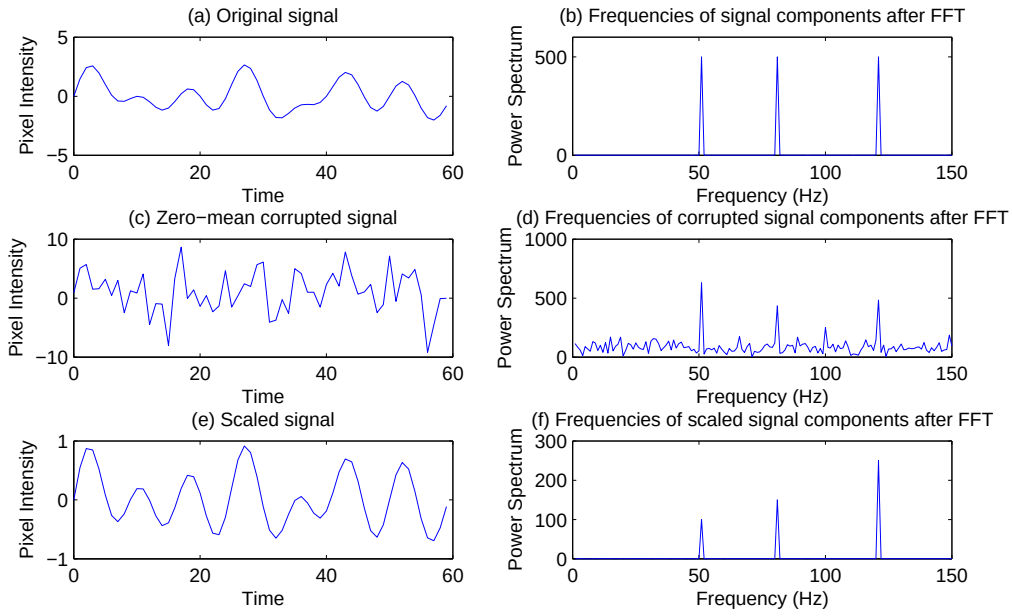


Figure 3.5: Frequency analysis.

Fig. 3.5 is an illustration of these properties of frequency analysis. Fig. 3.5(a) is the original signal represented by

$$f(t) = sin(2\pi \cdot 50t) + sin(2\pi \cdot 80t) + sin(2\pi \cdot 120t). \tag{3.3}$$

Since the signal consists of three components with frequencies $50Hz$, $80Hz$, and $120Hz$, it is impossible to tell all the three components from its time domain representation. However, after the Fast Fourier Transform, in Fig. 3.5(b), it is easily to tell the three components by their frequencies. Fig. 3.5(c) shows when the

original signal was added by zero-mean noise, the signal in time domain gets highly corrupted, and it is difficult to tell whether Fig. 3.5(a) and Fig.3.5(c) are related or not. However, after the FFT, in Fig. 3.5(d), the frequencies of its components remain unchanged, though their intensities have been affected by noise. Fig. 3.5(e) illustrates the scenario in which the three components of the signal get scaled by different factors, as in

$$f(t) = 0.2 \cdot sin(2\pi \cdot 50t) + 0.3 \cdot sin(2\pi \cdot 80t) + 0.5 \cdot sin(2\pi \cdot 120t). \tag{3.4}$$

In the time domain, not only it is impossible to distinguish these three components, but also hardly to tell how much they have been scaled. However, in Fig. 3.5(f), after the FFT, the frequencies remain the same, and it is easy to tell how much they have been scaled, since the power spectrum values of these three are proportional to the square of the scaling factors, i.e. $0.2^2 : 0.3^2 : 0.5^2 = 4 : 9 : 25$.

Frequency-based patterns

With these properties of frequency analysis, the frequency-based patterns can be designed to find correspondences between the pixels on the captured images and the pixels on the projected patterns in the frequency domain.

The patterns should meet the following requirements:

1. For each pixel position on the patterns, it should have a unique frequency, so that the frequency can be used to uniquely locate the position on the pattern. Basically, start from the top to the bottom, the frequency can be from $1Hz$ to $320Hz$, for an image with 320×320 resolution.

2. The number of patterns has to satisfy with the Nyquist-Shannon sampling theorem. According to this theorem, the sampling rate has to be larger than twice of the maximum frequency. The sampling rate is actually the number of patterns.

3. The frequency of each pixel position should be easily separated from noise. For the positions with low frequencies, they are more easily to get interfered and immersed into low frequency noises. Hence, the designed patterns should not have low frequencies.

The frequency-based patterns are divided and designed into two groups, horizontal ones and vertical ones. For an image with a resolution of 320×320, a number of $320 \times 320 = 102400$ different frequencies are needed. If the frequencies start from $1Hz$, the maximum frequency will be 102400. According to the Nyquist-Shannon sampling theorem, at least a number of 204800 patterns are needed, which will make the capturing procedure very time-consuming. To reduce the capturing time, the patterns are divided into vertical ones and horizontal ones, so that each group has at least 640 images. Because of the low frequency noise problem, and Zhu et al. [4] show that the low frequency noise are all within $5Hz$, the frequency-based patterns are designed with frequencies starting from $11Hz$ to $330Hz$. In this way, the low frequency components will be intentionally eliminated. Since now the maximum frequency is $330Hz$, at least 660 images are needed. Hence, 675 images are generated for vertical and horizontal patterns respectively.

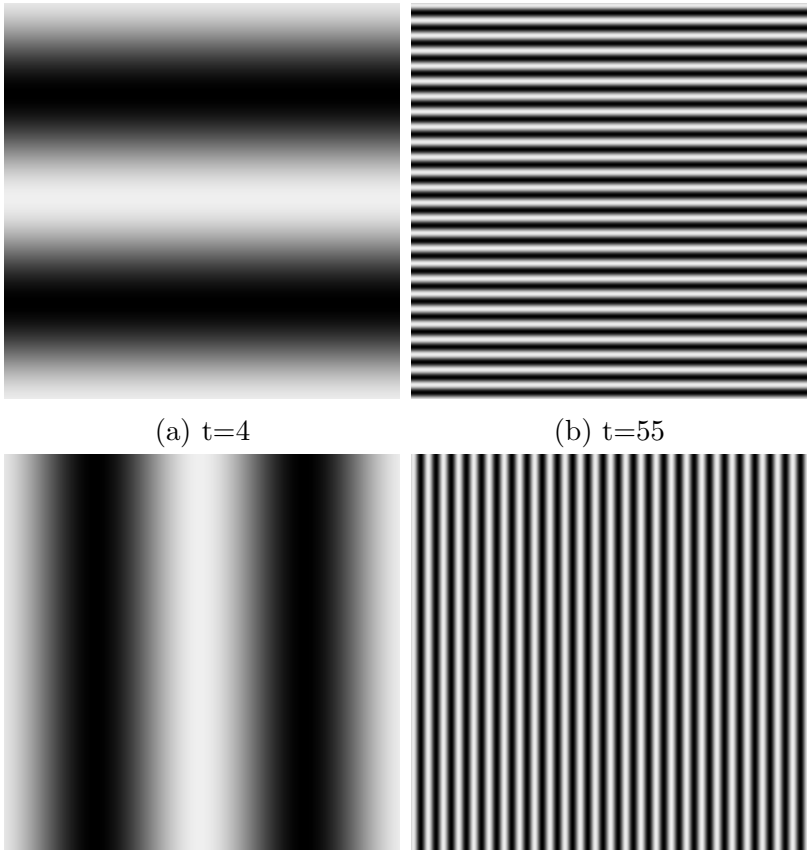
The patterns are designed as follows

$$I(i,t) = [cos(2\pi \cdot (i+10) \cdot t) + 1] \cdot 120 \tag{3.5}$$

where each pixel position on the patterns has an intensity $I(i,t)$, with variation of i and t. Here, t is the "time" index, and ranges from 0 to 1, with an interval of $1/675$, denoting 675 images in total. For the horizontal patterns, the intensities are the same for the same row, and different for different rows. In Eq. 3.5, i denotes the row index, and varies from 1 to 320, which means that the frequencies increases from the top row to the bottom row, and from $11Hz$ to $330Hz$. For the vertical patterns, it is very similar.

Since every object has a different volume, and sometimes when the object is very big, or has very detailed surface structures that need to be accurately reconstructed, the patterns must be redesigned with a higher resolution. Luckily, it is quite easy to increase the resolution of the pattern and cover a larger area using a similar equation as Eq. 3.5. In the experiments, projecting patterns with a resolution of 320×320 is found to be sufficient for objects reconstructed.

Fig. 3.6 illustrates some examples of the patterns.



(a) t=4 (b) t=55

(c) t=4 (d) t=55

Figure 3.6: Examples of the patterns designed.

Preprocessing

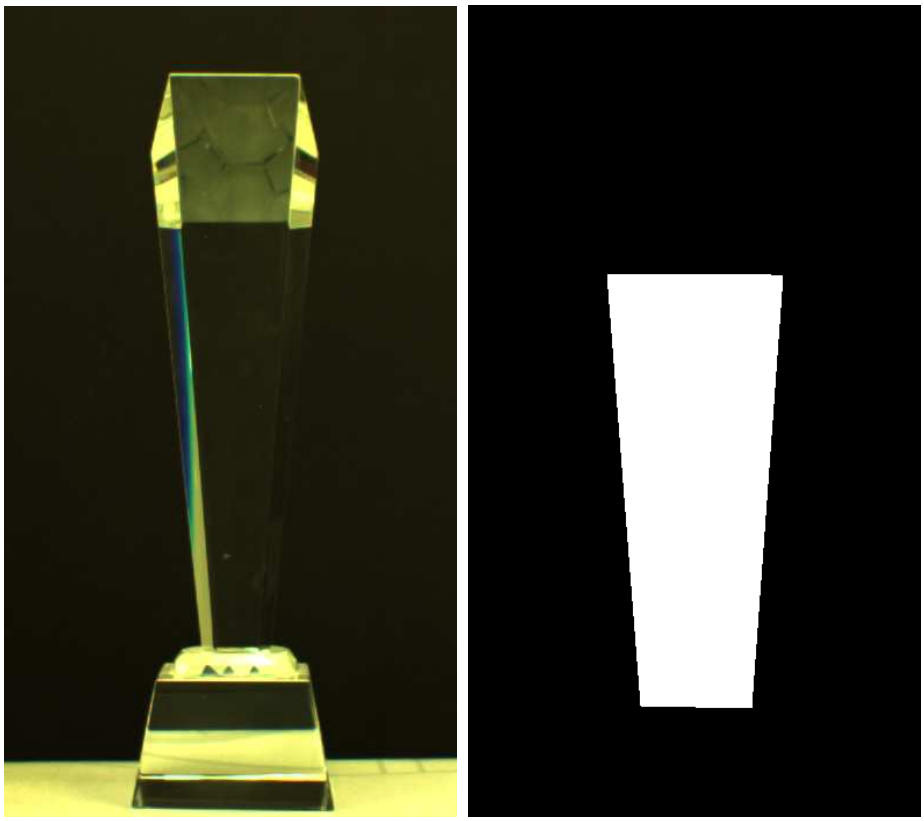
After using the projector to cast these patterns onto an object, a series of images can be acquired. But before analyzing the captured images, a preprocessing technique needs to be done to simplify the following steps.

In order to approximately define the regions of interest, a freely distributed software called GIMP2.8 [77] is used to extract the alpha matte. Here, the alpha matte is only used to define the region of the foreground object, so the alpha matte is a binary image. The region of interest is manually selected using the software, and the foreground pixels are assigned to have an intensity of 255 and the background pixels are set to be 0, as shown in Fig. 3.7.

Analyzing the captured images

After locating the region of interest, the captured images can be efficiently analyzed.

The key piece of information to utilize frequency is to convert a stack of im-



(a) (b)

Figure 3.7: (a) The original image of a scene, (b) The binary alpha matte.

ages into the frequency domain. Intuitively, since a stack of patterns are projected onto the object one after another, the patterns can be seen along the "time-axis." According to the frequency analysis description in section 3.2.2, for a certain pixel position of the patterns, the intensity varies periodically. Hence, when "seen" along the time-axis, these intensities can form a signal in the time domain, and the magnitudes of the discrete signal are the intensities. When transformed into the frequency domain, each signal has only one frequency, which uniquely defines its position (row or column) in the pattern.

Similarly to the frequency analysis of the patterns, the sequentially captured images can also be transformed into the frequency domain, except that the signal normally contains more than one frequencies, indicating that the pixel receives contributions from different positions of the patterns.

The challenge is to come up with a good data structure to "store" these captured images to reduce the analysis time. Fig. 3.8 is an illustration of our data structure for the captured images.

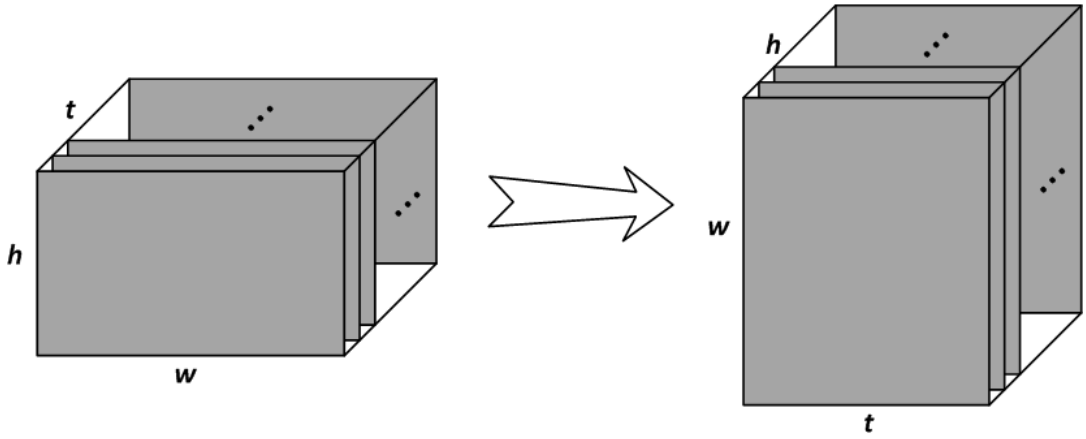


Figure 3.8: An illustration of our data structure for the captured images.

In Fig. 3.8, the left one shows that the captured images are "read" in one at a time and stored in a 3D matrix. Then, for each pixel position from left to right, from top to bottom, the intensity is read along the time-axis, and stored into the data structure shown on the right. In this way, the Discrete Fourier Transform can be applied one row at a time.

After storing these captured images, the whole data storage can be scanned and the DFT is conducted. For each pixel, the first step is to determine whether it is within the region of interest. The second step is to use the Discrete Fourier Transform to transform the signal from the time domain into the frequency domain.

Then the local maxima of the power spectrum are found in order to use the corresponding frequencies to locate the positions from which the original light paths originate. The reason to find the local maxima is because these peaks all have most of the contributions to the converged point and so they are all candidates of the first-order reflections. The reason of not choosing only the global maximum as the first-order reflection is because most of the times, since the object is transparent, the major portion of the light gets transmitted into the object, and only a small portion is reflected directly from the surface. Hence, the reflected light does not contain much energy, and so in the corresponding power spectrum, it is not the highest globally. However, comparing to the power spectra of other pixels in the neighbourhood, the first-order reflection can be at least locally maximum.

This is one of the main differences of the proposed method from the frequency-based environment matting method. In Zhu and Yang's method [4], they use a threshold to choose the components. If the threshold is very tight, they will only

choose frequencies with high energy. Since the monitor is used as the backdrop, a higher energy indicates a higher contribution, and the positions that give more contributions have a major effects to the results. However, for the proposed method, since the goal is to do 3D reconstruction of transparent and specular objects, the key is to find the correct correspondences, especially the correct correspondences of first-order reflections. Since the highest portion of contribution to the final image does not indicate that it is the first-order reflection. However, a first-order reflection is at least locally maximum, instead of a global maximum, of the power spectrum selected.

After finding the local maxima of the power spectrum, their frequencies can be correspondingly acquired. These frequencies uniquely locate a group of potential correspondences on the pattern. With the projector center, and the positions just found, the directions of the outgoing light from the projector can be computed. With the camera center, and the converged pixel, the direction of the incoming light to the camera can be obtained. Using linear triangulation [78, 79], the intersections of these light paths can be computed. These intersections are candidates for the point on the surface of the object, but only one of them is the correct point, which is the first-order reflection point. We select the first-order reflection point from these candidate points using a new labelling procedure.

3.2.3 Labelling

Fig. 3.9 is an illustration of multiple intersections. The converged pixel P_0 and the camera center C can define only one direction, $\vec{i}$, while the projector center Pj with the contributing pixels in the pattern can determine multiple directions, $\vec{o_1}$, $\vec{o_2}$, $\vec{o_3}$ and $\vec{o_4}$. These directions intersect along the incoming direction at P_1, P_2, P_3, and P_4.

Among these intersections, intuitively the one nearest to the camera center should be the first-order reflection point. However, there is an exceptional case. Although P_4 is nearer to the camera than P_1, it is not the first-order reflection point. Because as shown in Fig. 3.9, the direction $\vec{o_4}$ first refracts into the object, and after a few refractions and reflections, a part of the light gets into the camera through pixel P_0. However, in reality, this scenario is quite rare, and even when it happens, the contribution is so small that it does not satisfy the local maxi-

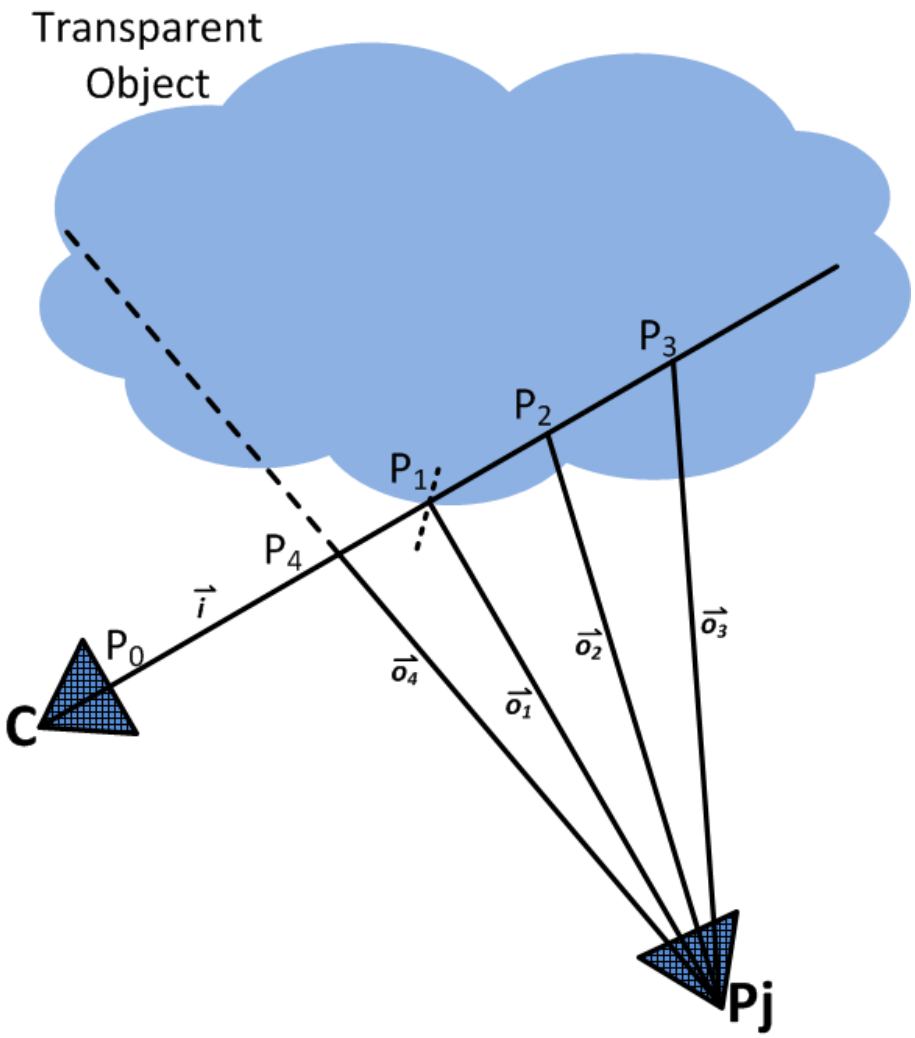


Figure 3.9: An illustration of multiple intersections.

mum selection criterion. Hence, normally this point P_4 is not chosen as one of the candidates.

Now that the conclusion is drawn that the first-order reflection point should be the nearest one to the camera center, a labelling method is used to select the point among all the candidates. We use a method inspired by Chen et al. [80] to do labelling.

Energy function

Generally, a labelling method is to label all the candidates, define an energy function based on their properties, and choose the labelling that can minimize the energy cost. Similar to [80], the energy function is defined based on Markov Random Field, also known as MRF, in

$$E(f_p) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q), \tag{3.6}$$

where p is the pixel within the region of interest in the captured image, f_p a label of pixel p, and $f_p \in \mathcal{L}$, where $\mathcal{L}$ denotes the label space. $D_p(f_p)$ denotes the data term, showing the cost of assigning label f_p to the pixel p. $\mathcal{P}$ is the pixel space of region of interest. $\mathcal{N}$ denotes the neighbor pixels of pixel p. $V_{p,q}(f_p, f_q)$ is the smoothness term, denoting the cost of assigning f_p to pixel p and assigning f_q to pixel q, which is the neighbor of pixel p. The details of the data term and the smoothness term are described below.

Data term

The data term is illustrated by the distance from the triangulated point to the camera center. Since the first-order reflection point is the closest triangulated point to the camera center, the data term illustrates this property. In the proposed method, the triangulation is first conducted and the intersections are obtained as candidates for the first-order reflection point. Then, since the 3D coordinates of intersections are in the camera coordinate system, it is easy to calculate the Euclidean distance from each intersection to the camera using

$$D_p(f_p) = \sqrt{\sum_{i=1}^{3} (f_{pi} - C_i)^2}, \tag{3.7}$$

where f_{pi} denotes the ith value of the 3D coordinates of the pixel p, after assigning the label f_p to it. C_i is the ith value of the 3D coordinates of the camera center. Since the camera center is the origin of the coordinate system, the coordinates of the camera center are actually (0,0,0). The data term for each pair of correspondences are calculated.

Smoothness term

In Eq. 3.6, $V_{p,q}(f_p, f_q)$ represents the smoothness term. Without loss of generality, it is assumed that the reconstructed object does not have sudden changes in shape, so that the smoothness property can be used. The 8-neighbors of the pixel on the captured image are chosen, the smoothness term of the pixel with each of its neighbor is calculated using

$$V_{p,q}(f_p, f_q) = |D_p(f_p) - D_q(f_q)|, \tag{3.8}$$

where $D_p(f_p)$ is the data term of label f_p, denoting the distance from the triangulated point to the camera center. The data term $D_q(f_q)$ is denoted for pixel q in a similar way. The pixel q is one of the neighbors of the pixel p in the captured image.

Minimizing the energy function

The energy function is minimized using the classical graph cuts. According to the results from [81], the expansion move algorithm introduced by [82] gives faster and better results than other methods in general. Hence, specifically, the expansion move algorithm of graph cuts is chosen to optimize the energy function.

3.2.4 Post-processing

After labelling, the first-order reflection points can be computed and assembled to reconstruct the surface of the object. However, normally the camera has a higher resolution than the projector, and the farther the pattern is cast, the wider each pixel from the pattern covers. Hence, we not only need to find correspondences from the camera to the projector, but also need to do the "reverse."

The previous steps have chosen the correspondence from the pixel in the pattern to the pixel in the captured image, but because of the resolution difference, the "one-to-one" correspondence is actually "many-to-one." In order to get a finer result, the "reverse" must be done to find the "one-to-one" correspondence from the camera plane to the projector plane. The implementation is quite easy. We just need to find, for each pixel in the pattern, which pixel in the camera plane corresponds to it and use the average position as the correct correspondence of the captured image.

This strategy has another advantage. After the frequency analysis, all the correspondences are integers, because they denote the pixel positions on the patterns and images. But in reality, one pixel may not be accurate to represent an exact 3D point on the object. If the pixel on the pattern covers more than one point on the surface of the object, and we use this pixel to do the triangulation, we may have the same convergence point for all of these points, and that will be inaccurate. Hence, we need floating point numbers to increase the accuracy. After the "reverse" step, the correspondences in the camera image are floating point numbers, which are more

accurate.

There are more strategies to do post-processing. The basic idea is to get result from previous steps first, observe it, and find a way to refine the result.

3.2.5 Reconstructing the surface

Since the first-order reflection points are denoted as $3D$ coordinates, we can use the open source system called MeshLab [83] to reconstruct the surface of an object. The input for MeshLab is point cloud. Since our result is in the format of point cloud, we can use MeshLab to open it and transform it into a mesh if necessary.

3.3 Summary

In this section, a novel method is introduced to do the frequency-based 3D reconstruction for transparent and specular objects. In brief, the frequency-based environment matting method is adapted to identify the correspondences between pixels in the captured images and pixels in the patterns. Then, the linear triangulation is conducted to find the convergence points from the correspondences. Since only one of these points can be the first-order reflection point, a labelling method is introduced to select the real one. For labelling, the Markov Random Field is used to define the data term and the smoothness term, based on the properties of the object and the first-order reflection. Then the expansion move algorithm of graph cuts is applied to minimize the energy function of the MRF, and obtain the nearest one to the camera center as the first-order reflection point. Since the projector and the camera have different resolutions, and also during the experiment there may be some noise or error, a few post-processing techniques are proposed to refine the result. In the end, MeshLab is used to illustrate the surface of the object.

Chapter 4

Experiments and Results

In this chapter, the experimental steps are summarized and a few detailed techniques are discussed to solve some problems that are encountered during experiments. The experimental results are presented using different objects.

4.1 Design of experiments

The experimental setup is shown in Fig. 3.3, which is quite similar to the traditional setup of the structured light method.

The experimental steps are quite easy. The first step is to do calibration of the camera and the projector. We used the method introduced by Falcao et al. [84]. Their method is an extension for Bouguet's Camera Calibration Toolbox [85]. First, the traditional camera calibration using Bouguet's Camera Calibration Toolbox is conducted. Second, the projector calibration is carried out by treating the projector as an inverse camera. After calibration, an image with a solid white pattern is captured and the alpha matte is extracted manually. Then we sequentially project horizontal and vertical patterns onto the object and capture the images with a calibrated camera. With the acquired images, we use the Discrete Fourier Transform to convert the pixel signal into the frequency domain. Then, we find the dominant frequencies using some thresholds. The frequencies correspond to pixel positions on the pattern. In this way, we find correspondences between pixels in the captured images and pixels in the projected patterns. After finding the correspondences, we use triangulation to find candidate points for first-order reflection. In order to select the correct first-order reflection points, we use the distance from the point to the camera center as the data term in the Markov Random Field, and use an extension of the graph cut algorithm to minimize the energy function. The selected points

after optimization are the first-order reflection points on the surface of the object.

Because of the differences in resolution between the camera and the projector, some post-processing techniques are used to refine the experimental results. One technique is to use a "window", and get the average of the z values. If the z value of one pixel within this window is far from the average, we regard it as an outlier and remove it. Here, a threshold will be helpful. The reason to use only the z value, instead of using the distance to the camera center, is because the pixels within a small area have similar z values, and using the z value is much simpler than using the distance value. However, when the z value does not work, we can use the distance instead. Another technique to do post-processing is to find the correspondences from the projector to the camera, and repeat the triangulation again. In this way, only one corresponding point is acquired for each pixel, and this point is regarded as the first-order reflection point.

After getting the surface points in the form of a point cloud, the surface of the object can be reconstructed using MeshLab.

To sum up, both the capturing procedure and the image processing procedure are accordingly described in Algorithm 1 and 2.

Algorithm 1 Pseudocode for Image Capturing Procedure

/*——**Images for Calibration**——*/
Project the checkerboard pattern onto the right side of a white board with another checkerboard printed on its left side
Move the white board to different positions and at each position capture the projected pattern along with the planar pattern on the white board at the same time

/*——**Images for Preprocessing**——*/
Display a solid white pattern onto the object and capture the scene

/*——**Frequency-Based Patterns**——*/
Display the sequence of horizontal patterns onto the object, and for each pattern capture the image as H_i
Display the sequence of vertical patterns onto the object, and for each pattern capture the image as V_i

Algorithm 2 Pseudocode for Image Analyzing Procedure

```
/*——Calibration——*/
Use Bouguet's Camera Calibration Toolbox to do the camera calibration
Use Falcao et al.'s method to do the projector calibration

/*——Preprocessing——*/
Use GIMP2.8 to manually extract the regions of interest from the captured image

/*——Frequency Analysis——*/
Read in all the H_i and V_i, store them in the form of a 3D matrix, and transform it into matrix M_h and M_v, so that each row of the new matrix is a 1D array containing the intensities of a certain pixel position through captured images
for every image pixel position C(x, y) do
  if α(x, y) ≠ 0 then
    Perform the Discrete Fourier Transform to obtain M_h(x, y) and M_v(x, y)
    Use horizontal and vertical thresholds to select dominant local maximum ones, and find their corresponding frequencies
    For each frequency, locate the corresponding pixel on the patterns
    Do the linear triangulation, so that for each image pixel, find multiple candidates on the object
  end if
end for

/*——Labelling——*/
Use the triangulation result to get the distances from points to the camera center, and denote them as data terms
Define the energy function based on the Markov Random Field formulation
Use the expansion move algorithm of Graph Cuts to minimize the energy function and find the optimal one as the chosen label for the first-order reflection point

/*——Post-Processing——*/
Use the z value of the 3D point, and use a shifting window to refine the results (see Section 3.2.4)
Convert the correspondences from "image to projector" to "projector to image"
Calculate the mean position of the pixels in image from the corresponding pixel in projector
Do triangulation again for the new one-to-one correspondences

/*——Reconstructing the Surface——*/
Use the point clouds and MeshLab to reconstruct the surface of the object
```

There are some techniques that need to be mentioned here in order to get good experimental results. The first one is that the relative positions between the camera, the projector and the object need to be adjusted before projecting the solid white pattern and the frequency-based patterns to the object. Since the method uses the

reflection from the object to reconstruct its surface, and normally the reflection is not strong or is interfered by other light effects, it is important for the camera to receive as much incoming reflected light from the object as possible. For different objects, because their shapes and surface properties vary, different relative positions are required accordingly.

Another technique is that some thresholds are needed in order to select the dominant frequencies. As stated in Chapter 3, the local maxima of the frequencies denote the candidates for the first-order reflection points. However, because of the noise and errors when capturing the images, many local maxima are actually not the results of the reflections from different layers of the object, especially the frequencies with very low energy, as shown in Fig. 4.1. Hence, it is important to use thresholds to select the valid candidates. Take images with horizontal patterns as an example, after using the Discrete Fourier Transform to convert these captured images into the frequency domain, for each pixel in the captured images, multiple frequencies that are local maxima can be obtained, for example, in Fig. 4.1. Because of the symmetry property of the transform, only the first half of the frequencies, i.e. the lower frequencies shown in Fig. 4.1, are valid for the candidate selection. Intuitively, there are two major local maxima. Additionally, there are more local maxima with very low energy and they are not the valid candidates. Hence, a threshold is used to select the local maxima with a relatively higher energy. The technique to choose the threshold is quite simple. We can randomly choose a pixel in the region of interest from the captured images, and calculate its spectrum. Normally the first 10 local maxima can ensure to include the frequency corresponding to the first-order reflection point. Based on the distribution of the energy of its frequencies, we can estimate a value for the threshold that includes the first 10 local maxima. Using the result after the frequency analysis, but before the labelling procedure, to do the linear triangulation. If the threshold does not include the whole surface of the reconstructed region, the value of the threshold is reduced. Since the images are captured with horizontal and vertical patterns, two thresholds corresponding to the two patterns are needed.

4.2 Experiments

In this section, detailed experimental results with different objects are illustrated.

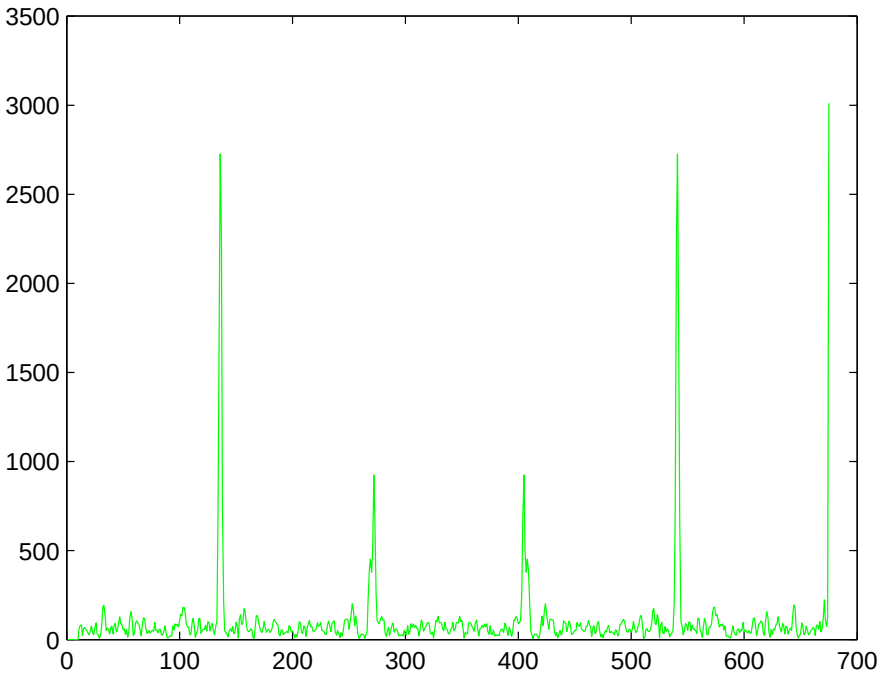


Figure 4.1: An example of frequency analysis for the trophy with multiple faces at pixel (i=432, j=321).

As a comparison, the classical Gray code method is used to do 3D reconstruction and the results are accordingly compared with those using the proposed method. The reason to use the Gray code method is because it has been widely used as a comparison with other 3D reconstruction methods using structured light. The results of using the Gray code method are quite good for opaque objects. Another comparison is with the ground truth. To obtain the ground truth of a transparent object, a cosmetic face powder is mixed with water as "paint" and gently brushed onto the object. After a few minutes, the paint will dry and the Gray code method is used to reconstruct the "opaque" object and the result is used as the ground truth. The reason to use cosmetic powder is that the paint needs to be opaque and can be easily washed off, so that the experiments can be repeated. In addition, it should have no chemical reaction with the object. However, when the object has detailed structures, the paint may occlude such features, in which case, only the picture of the object is used as the ground truth for qualitative evaluation.

In order to embed the Gray code method into the experiments to get results for comparison and the ground truth results, and avoid disturbing the experimental

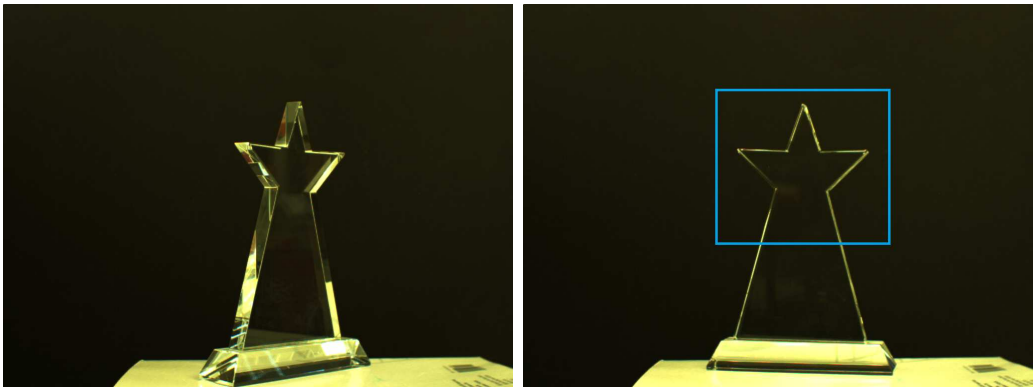
setup, the whole experiment is designed as follows.

1. Adjust the setup so that the camera can receive as much reflected light as possible.
2. Do the camera-projector calibration.
3. Put the object in the scene and adjust its, and only its, position, so that the camera can receive enough reflection from the object. Do not touch the camera, nor the projector, since they have been calibrated.
4. Project the solid white pattern to the scene with the object and capture the image for alpha matte.
5. Adjust the shutter speed of the camera in *FlyCap2*, and project the frequency-based patterns onto the object. Capture the scene with the camera simultaneously with each horizontal pattern and each vertical pattern.
6. Use Gray code patterns and get 20 vertical images and 20 horizontal images.
7. Use the solid black pattern and the solid white pattern to get reference images for the Gray code method.
8. Brush the paint onto the object gently, without moving the object.
9. Use Gray code patterns to get 40 images of the scene, and use the black and white patterns to get reference images.
10. Analyze all three sets of images with corresponding methods and use MeshLab to illustrate the results.

4.2.1 Qualitative results

The objects used for the experiments include a star trophy, a cone trophy with multiple faces, a big vase, a small vase, an anisotropic metal cup, a plastic cup with two layers and a plastic bottle with green dishwashing liquid in it.

Star trophy



(a) A star trophy (b) The reconstructed region

Figure 4.2: The star trophy and its reconstructed region.

As shown in Fig. 4.2a and Fig. 4.2b, the star trophy is solid and totally transparent with no inner structure. When the patterns are projected to the object, most of the light goes through it and gets reflected by the background. The reflection from the surface of the object is interfered by the reflection from the back of a surface and also by the reflection from the background. In addition, because the object has sharp edges, the highlight is strong and cannot be avoided, and also can interfere with selecting candidates for the first-order reflection. The traditional methods using structured light fail because of the highly optical interactions, shown in Fig. 4.13. However, using the proposed method, good results can be acquired (Fig. 4.11).

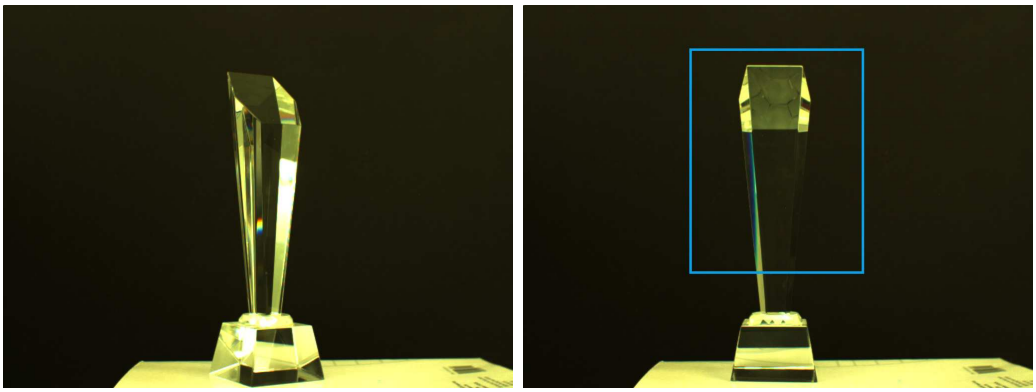
Fig. 4.11 shows the reconstruction results using the proposed method, compared with the ground truth. As shown in Fig. 4.11(a)(b)(c), the surface of the object is reconstructed smoothly. However, there are a few small holes in the results, such as the lower right one at the "corner" of the object. The reason for these flaws is because of the highlight. The proposed method fails when the highlight is strong. For pixels with strong highlight, their intensities have little variations. Hence, when transform the series of intensities into the frequency domain, the energy of the corresponding frequency can be as low as noise. Hence, the pixels in the highlight region may get wrong or no correspondence.

Fig. 4.12 shows the results after using frequency analysis of the proposed method, but before the labelling procedure. Multiple "surface" layers are observed, which correspond to the candidates for the first-order reflection points. After doing linear triangulation with these correspondences, wrongly reconstructed surfaces, as well as

the correct surface, are obtained. Fig. 4.12 shows the importance of the labelling procedure. Noticing that in Fig. 4.12c, only the upper left part of the object was wrongly reconstructed and has multiple layers. The reason is that the first-order reflection from this part of the object is highly interfered by the reflection from the background. However, even with these wrong initial candidates, the correct correspondences are also selected using the labelling procedure. Hence, this result shows that using the labelling method, the first-order reflection can be detected and the correct surface can be reconstructed (Fig. 4.11).

Fig. 4.13 shows the reconstruction results using the structured light method with the Gray code patterns. As discussed before, because of the optical interactions, the first-order reflection is difficult to detect. Since the Gray code method uses the intensities directly in the time domain, many errors occurred during the correspondence process. The upper left part of the object is wrongly reconstructed, which illustrates the wrong correspondences. Comparing with the ground truth shown in Fig. 4.11(d)(e)(f), the proposed method can acquire much better results than the Gray code method.

Cone trophy with multiple faces



(a) A cone trophy with multiple faces (b) The reconstructed region

Figure 4.3: The cone trophy with multiple faces and its reconstructed region.

Fig. 4.3b shows the area of the object that is reconstructed, which is the largest face of the object. Since the top face is tilted, it refracts a large amount of light into the object, and this portion of light, together with other lights, interacts with the first-order reflection from the largest face, making it quite difficult to do 3D reconstruction using the Gray code method. Shown in Fig. 4.16, the big hole in

the middle is where the refraction from the top face refracts out of the object. Although the proposed method is also affected by the strong refractions, shown in Fig. 4.14(a)(b)(c), the new method turns out to be much better than directly using the intensity. In Fig. 4.14(a)(b)(c), the holes are much smaller and the surface looks much smoother than that using the Gray code method (Fig. 4.16).

Similar to the results of the star trophy, the frequency analysis results (Fig. 4.15) before the labelling procedure also show multiple layers of "surfaces," which are also because of the multiple candidates for the triangulation. Using labelling, the first-order reflection points can be mostly acquired. Although for some parts, a few holes can be observed, the results of reconstruction using the proposed method are better than that using the Gray code method.

Big vase

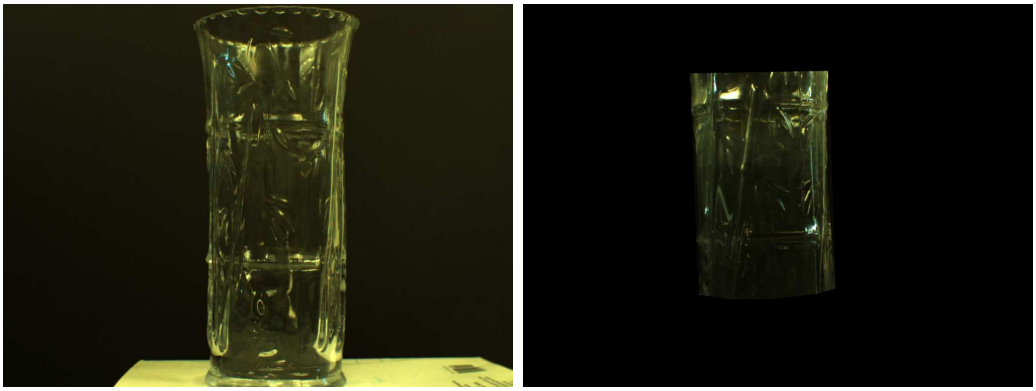


Figure 4.4: The big vase and its reconstructed region (i.e. the ground truth).

The big vase has decorative bamboos and leaves patterns engraved all around its surface, and these detailed structures have an active interaction with light. Strong highlights from these structures can be observed. Because of the complex surface structures, it is very hard to find an appropriate setup so that the camera can receive most of the surface reflections. Hence, the lack of captured reflection is a big challenge for the 3D reconstruction of the big vase.

Fig. 4.4 shows the big vase and its reconstructed region. Since the object has a detailed surface structure, when using the cosmetic powder mixed with water and brushing the "paint" onto the surface, the details are covered. Hence, the image captured by the camera is used as the ground truth for this object. Fig. 4.17 shows

the results using the proposed method. Although there are many holes because of the lack of the received reflections and of the highlights, the features of bamboo and leaves can still be observed. Although the proposed method did not reconstruct the part around the protrusions, it did acquire some details of the patterns. Fig. 4.18 shows the results before labelling. The highlights introduced many wrong correspondences, leading the linear triangulation results to be poor. For the Gray code results shown in Fig. 4.19, other than the trunk of the bamboo, it fails in reconstructing the surface.

Small vase

Figure 4.5: The small vase and its reconstructed region (i.e. the ground truth).

Fig. 4.5 shows the small vase and its reconstructed region. This small vase has a surface like a pineapple, and every protrusion is similar to each other except for the size. Not only highlights are easily observed, there are also dispersions, complex reflections and complex refractions. The light paths are untractable. Using the Gray

code method will totally fail for this object, as shown in Fig. 4.22. Because of the complex interactions, it is difficult to acquire light-paths and to use refraction-based methods. Because the refraction, reflection and dispersion will highly affect the first-order reflection, making its intensity indistinguishable from other light effects. The traditional reflection-based methods designed specifically for opaque objects, such as the Gray code method, will most likely fail.

However, though the intensities are hard to be detected, let alone be transformed into frequency domain, the first-order reflections invisible to the human eye can actually be transformed and the frequencies are more distinct than their corresponding intensities. Fig. 4.20 shows the result using the proposed method. A lot of detailed structures are reconstructed. Although there are still wrong points because of the highlights, most of the reconstructed 3D points illustrate the expected features of the surface of the object, such as the "pineapple" texture.

Metal cup

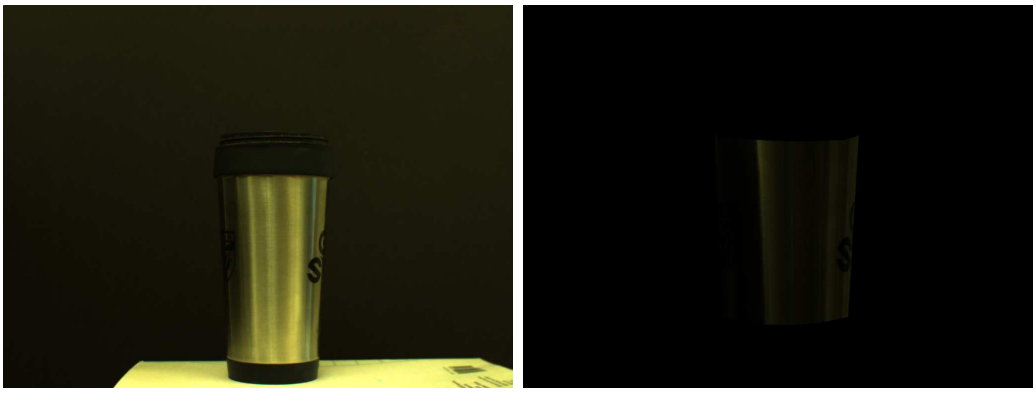


Figure 4.6: The metal cup and its reconstructed region (i.e. the ground truth).

The proposed method has a large range of applicable objects. Not only for traditional opaque objects, and transparent objects, but also for specular objects with anisotropic surfaces, the method can produce quite acceptable results. Reconstructing objects with anisotropic surface is very challenging, because methods using normal Lambertian reflectance or specular reflections tend to fail on anisotropic surfaces. Fig. 4.6 shows a metal cup with an anisotropic surface and its reconstructed region. Fig. 4.25 shows the results using the Gray code method, we can easily observe the holes in the middle and at the sides. That is because the surface reflects

light anisotropically, and the intensities of these reflections are wrongly interpreted when finding the correspondences. That is the reason why in Fig. 4.25c, wrongly triangulated points can be observed in the front and at the back of the reconstructed surface. Fig. 4.23 shows the reconstructed results using the proposed method. Comparing to the results of the Gray code method, our results have much smaller holes and smoother reconstructed surface. Fig. 4.24 shows the results before labelling, with all candidates for fist-order reflection. Using the labelling procedure, most of the incorrect candidates can be detected and eliminated.

Plastic cup

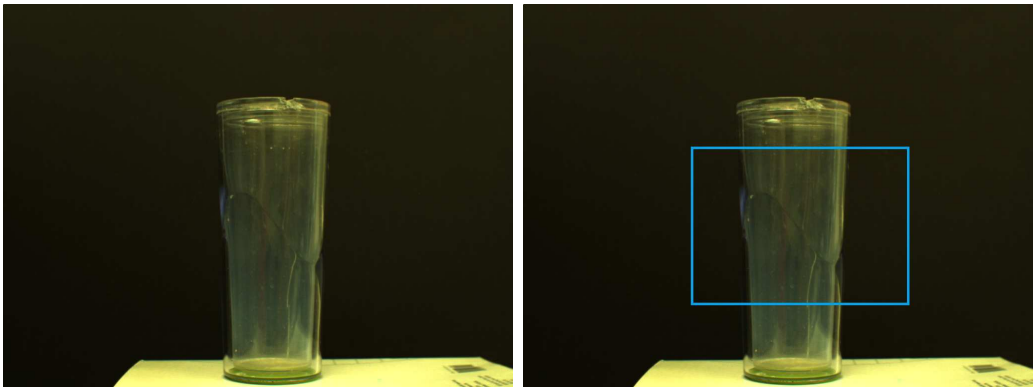


Figure 4.7: The plastic cup with two layers and its reconstructed region.

The plastic cup shown in Fig. 4.7 is quite challenging to be reconstructed because it has two layers. The second layer (the inner one) has strong reflections, and the reflections are quite close to that from the first layer. Since the frequencies after the Discrete Fourier Transformation are also quite similar, it is very hard to detect the real first-order reflections. The results shown in Fig. 4.26 is not very good. Although the reconstructed surface is smooth and the detailed "wave" of the surface is preserved, big holes can be observed. Comparing with the results before the labelling procedure (Fig. 4.30), the big holes come from wrongly detected correspondences. However, comparing to Fig. 4.31, the proposed method has a much better result.

Bottle with dishwashing liquid

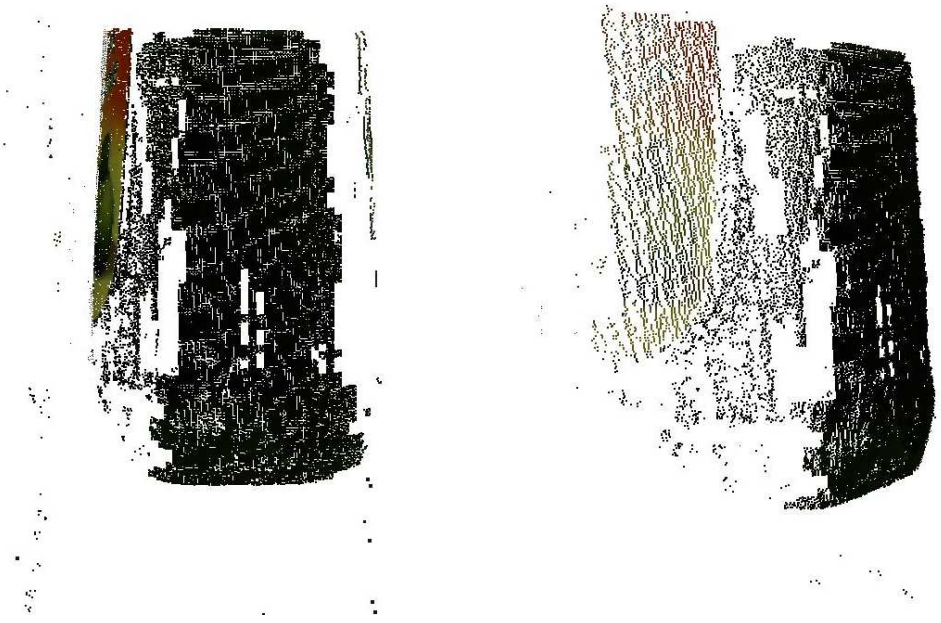
Figure 4.8: The bottle with dishwashing liquid and its reconstructed region.

Fig. 4.8 shows a plastic bottle with a green dishwashing liquid inside. The dishwashing liquid is transparent and since it has a different refraction index from the plastic bottle, refraction and reflection happen at the interface between the bottle and the dishwashing liquid. Fig. 4.9 shows the reconstruction results using the proposed method. The holes in the middle indicate the points there received too strong highlight to be reconstructed. The holes on both sides of the object are due to the high curvature on the surface and the camera did not receive enough reflections from this part. The proposed method has a better result than that using the Gray code method, as shown in Fig. 4.31.

4.2.2 Quantitative results

For the quantitative results, two very challenging objects are used. The first one is the star trophy, as shown in Fig. 4.2. The second one is the cone trophy (Fig. 4.3), with multiple faces. The results are compared with the ground truth. As a comparison, the results of the Gray code method are also compared with the ground truth. According to Chapter 3, the pixels in the patterns correspond to the averaged correspondences, which are in floating point format, in the captured image. Hence, the corresponding pixels in the captured images are compared to the same corresponding points in the patterns of our method and in the ground truth.

Eq. 4.1 defines the root mean square (RMS) error of the correspondences of the results of the frequency-based reconstruction method. (x, y) denotes the point in the patterns that has a corresponding pixel in the captured images, and the

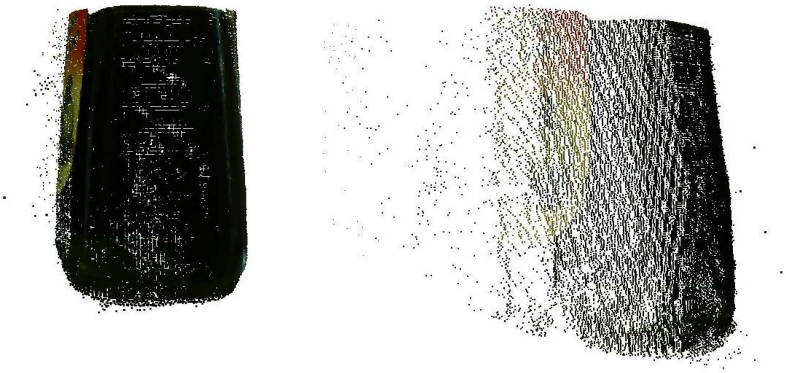


(a) Front

(b) Left side

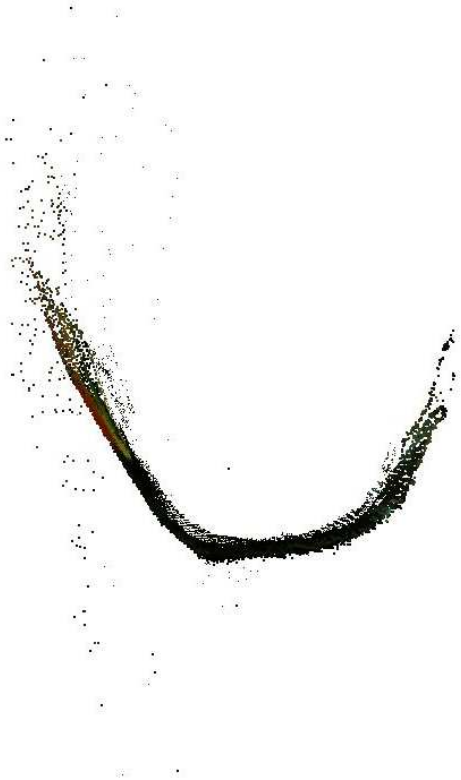
(c) Top

Figure 4.9: Reconstruction results for the plastic bottle with a green dishwashing liquid using our method.



(a) Front

(b) Left side



(c) Top

Figure 4.10: Ground truth for the plastic bottle with a green dishwashing liquid.

corresponding pixel is in floating point format and is within the region of interest. $C_F(x, y)$ denotes the pixel in the captured images, which corresponds to (x, y) in the patterns using the frequency-based reconstruction method. $C_T(x, y)$ denotes the pixel in the captured image of the ground truth, which corresponds to (x, y) in the patterns. N_F is the number of the points (x, y) which have corresponding pixels within the region of interest in the captured images. The root mean square error of correspondences for the Gray code is similar to that of the frequency-based reconstruction method and is defined in Eq. 4.2. $C_G(x, y)$ denotes the corresponding pixel in the captured images acquired by the Gray code method, which is also in the region of interest. N_G is similar to N_F, which is the number of the points (x, y) which have corresponding pixels within the region of interest in the captured images using the Gray code method.

According to Eq. 4.1 and Eq. 4.2, only corresponding pixels within the region of interest are compared to the ground truth. For pixels that are outside of the region of interest, since there is no corresponding pixel in the ground truth to be compared with, they are simply eliminated from the comparison.

In addition to the RMS errors, another way to illustrate the quantitative results is shown in Eq. 4.3 and Eq. 4.4, which are used to show the "score" of the frequency-based method and the Gray code method. In Eq. 4.3, N_F means the number of the corresponding pixels within the region of interest using the frequency-based method. N_{all} denotes the total number of the corresponding pixels within the region of interest of the ground truth. $\frac{N_F}{N_{all}}$ denotes the fraction of the correspondences within the region of interest to be reconstructed by the frequency-based method. The bigger the value of $\frac{N_F}{N_{all}}$, the higher the reconstructed resolution of the results. $C_{F\_RMS\_error}$ denotes the RMS error of the correspondences within the region of interest using the frequency-based method compared with that of the ground truth. The smaller the value of $C_{F\_RMS\_error}$, the better the result. The combination of $\frac{N_F}{N_{all}}$ and $C_{F\_RMS\_error}$, which is $Score_{F\_correspondences}$, illustrates the results of the correspondences using our method compared with that of the ground truth, with consideration of the resolution. The higher the value of $Score_{F\_correspondences}$, the better the result. Eq. 4.4 shows the results of the Gray code method compared with ground truth, with consideration of the resolution of the correspondences.

$$C_{F\_RMS\_error} = \sqrt{\frac{1}{N_F} \sum_{x=1,y=1}^{x=1024,y=768} \Big(C_F(x,y) - C_T(x,y)\Big)^2} \tag{4.1}$$

$$C_{G\_RMS\_error} = \sqrt{\frac{1}{N_G} \sum_{x=1,y=1}^{x=1024,y=768} \Big(C_G(x,y) - C_T(x,y)\Big)^2} \tag{4.2}$$

$$Score_{F\_correspondences} = \frac{\frac{N_F}{N_{all}}}{C_{F\_RMS\_error}} \tag{4.3}$$

$$Score_{G\_correspondences} = \frac{\frac{N_G}{N_{all}}}{C_{G\_RMS\_error}} \tag{4.4}$$

In addition to the comparison of the correspondences, the distances from the reconstructed points to the camera center are also compared to illustrate the results of the frequency-based method and the Gray code method.

Eq. 4.5 defines the root mean square error of the distances for the results of the frequency-based reconstruction method. (x, y) denotes the point in the pattern that has a corresponding pixel within the region of interest in the captured image. $D_F(x, y)$ denotes the distance from the reconstructed point on the surface to the camera center using the frequency-based reconstruction method. $D_T(x, y)$ denotes the distance from the reconstructed point of the ground truth on the surface to the camera center. N_F is the number of the compared distances. The RMS error of the distances for the results of the Gray code method is similar to that of the new method and is defined in Eq. 4.6. $D_G(x, y)$ denotes the distance acquired by the Gray code method. N_G denotes the number of reconstructed points that have corresponding pixels within the region of interest using the Gray code method.

Eq. 4.7 and Eq. 4.8 are similar to Eq. 4.3 and Eq. 4.4.

$$D_{F\_RMS\_error} = \sqrt{\frac{1}{N_F} \sum_{x=1,y=1}^{x=1024,y=768} \Big(D_F(x,y) - D_T(x,y)\Big)^2} \tag{4.5}$$

$$D_{G\_RMS\_error} = \sqrt{\frac{1}{N_G} \sum_{x=1,y=1}^{x=1024,y=768} \Big(D_G(x,y) - D_T(x,y)\Big)^2} \tag{4.6}$$

$$Score_{F\_distances} = \frac{\frac{N_F}{N_{all}}}{D_{F\_RMS\_error}} \tag{4.7}$$

$$Score_{G\_distances} = \frac{\frac{N_G}{N_{all}}}{D_{G\_RMS\_error}} \tag{4.8}$$

Table 4.1 shows the comparison results of the star trophy reconstruction results using the frequency-based reconstruction method and the Gray code method with that of the ground truth. Although the Gray code method has a higher resolution of the reconstruction result, it has a much higher RMS error than that of using the frequency-based method. For the frequency-based method, the RMS error is not as small as expected. The reason is because for the edge of the object, strong highlight makes the reconstruction incorrect for this part. For the holes with no 3D information in the reconstruction results, no comparison is made and these holes are neglected. For the wrongly reconstructed points using the Gray code method, since the ground truth does not have corresponding pixels for them to compare with, these corresponding pixels are also neglected. The score based on the correspondences and the score based on the distances show that the results of our method are much better than that of the Gray code method.

Table 4.1: The comparison results for star trophy reconstruction using the frequency-based reconstruction method and the Gray code method

	Frequency-based method	Gray code method
Number of reconstructed points within the region of interest	15580	16301
RMS error of the correspondences	4.0570	17.0101
Score based on correspondences	0.1130	0.0282
RMS error of the distances from the reconstructed points to the camera center	102.1415	856.9114
Score based on distances	0.0045	5.5991×10^{-4}

Table 4.2 shows the quantitative results of the cone trophy reconstruction using the frequency-based method and the Gray code method compared with the ground truth. For the Gray code method, only a small part in the middle failed to do the reconstruction. Hence, the RMS error of the correspondences and the distances

are quite close to the results of the frequency-based method. Noticing that the new method reconstructs more points than the Gray code method for this object. The strategies to handle the holes and errors of the reconstruction are the same as for the star trophy reconstruction results. The score based on correspondences and the score based on the distances show that the reconstruction results of the frequency-based method are better than that of the Gray code method.

Table 4.2: The comparison results for cone trophy reconstruction using the frequency-based reconstruction method and the Gray code method

	Frequency-based method	Gray code method
Number of reconstructed points	26799	21900
The RMS error of the correspondences	3.2900	4.0794
Score based on correspondences	0.2683	0.1768
The RMS error of the distances from the reconstructed points to the camera center	45.9847	43.2997
Score based on distances	0.0192	0.0167

4.3 Conclusions

In this chapter, the setup and basic steps of frequency-based 3D reconstruction method of transparent and specular objects are presented. The algorithms are discussed and a 10-step outline is designed for doing the experiments using the proposed method, the compared method, and the ground truth. For the experiments, seven objects are reconstructed. Although some results are not as good as expected, the proposed method generally has better results than that using the Gray code method. The frequency-based reconstruction method has the following advantages.

1. It has a wide range of applicable objects. Since the proposed method is reflection-based, it can be used generally for all opaque objects, though it is not the most efficient method. In addition, the method can reconstruct not only Lambertian surfaces, but also some anisotropic surfaces, which has seldom been achieved before. Most importantly, the proposed method can

tackle the problem of optically active interaction of object with light, and reconstruct the surface of transparent objects with complex interior and exterior structures.

2. This method is quite robust to multiple reflections and refractions, and preserves detailed structures of the surface. No matter how complex the reflections and refractions may be, in most cases, the correct correspondences can be acquired using frequency analysis. With the labelling procedure, finding the first-order reflection is quite easy.

3. The patterns used are quite easy to generate, and can be adjusted to have higher or lower resolution. Hence, the method can do the reconstruction both effectively and efficiently.

4. The setup is quite simple and cheap. Only one normal camera and one ordinary projector are needed. There is no need to move the setup during the experiment. The calibration method is quite straightforward. In addition, since no poisonous material is needed to spray or paint on the object, the experiment is quite safe for researchers and environmentally friendly.

The shortcomings of the method are itemized as follows.

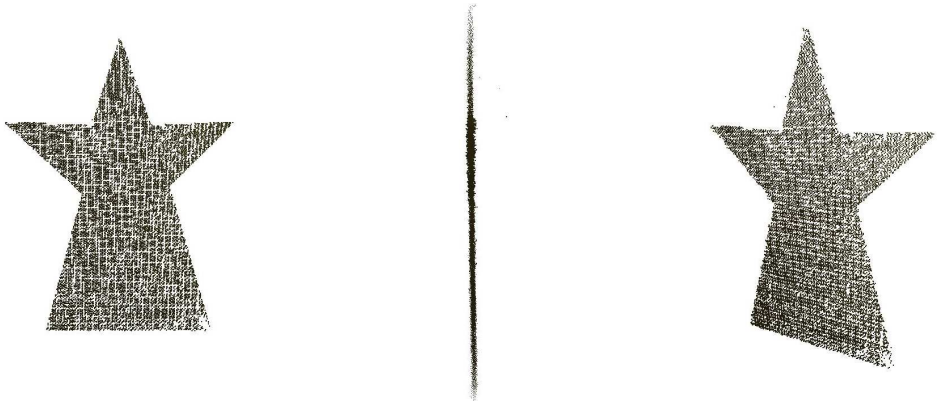
1. Since the reflection from the object is used, the relative position between the object, the camera and the projector is very important. In each experiment, their positions need to be adjusted so that the camera can receive as much reflection as possible.

2. Since the camera and the projector are on the same side of the object, the reconstructed region is limited. When the object has self-obstructions, the region will be smaller. Hence, a turntable is needed to reconstruct the whole object.

3. When the interior has a strong reflection structure, and the structure is too close to the surface of the object, this method will fail because it is difficult to distinguish two close surfaces. However, if the interior structure is a little farther from the surface, the correct outer surface can still be reconstructed.

4. The frequency-based reconstruction method cannot fully tackle the problem with highlights. Although the surface with detailed structures can still be reconstructed to some extent, with interference of the highlights, the surface is not fully reconstructed.

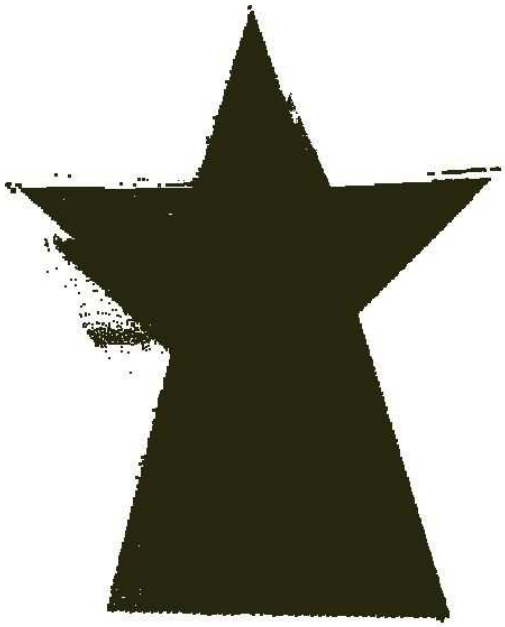
5. For the surface structure with high curvatures, this method can fail because of the limitation for the received reflections.



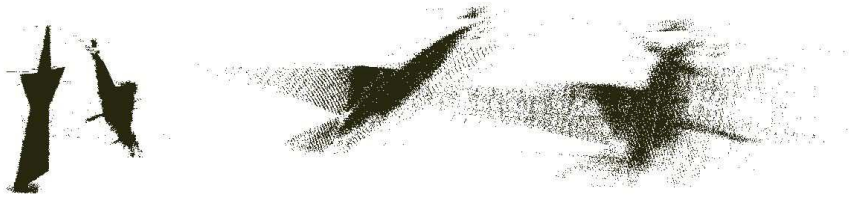
(a) Front (b) Left side (c) Right side

(d) Front (e) Left side (f) Right side

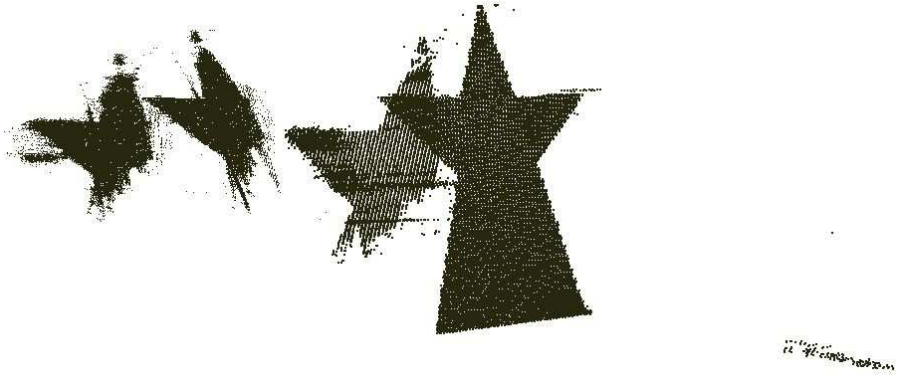
Figure 4.11: Reconstruction results for star trophy using our method ((a)(b)(c)), compared with the ground truth ((d)(e)(f)).



(a) Front



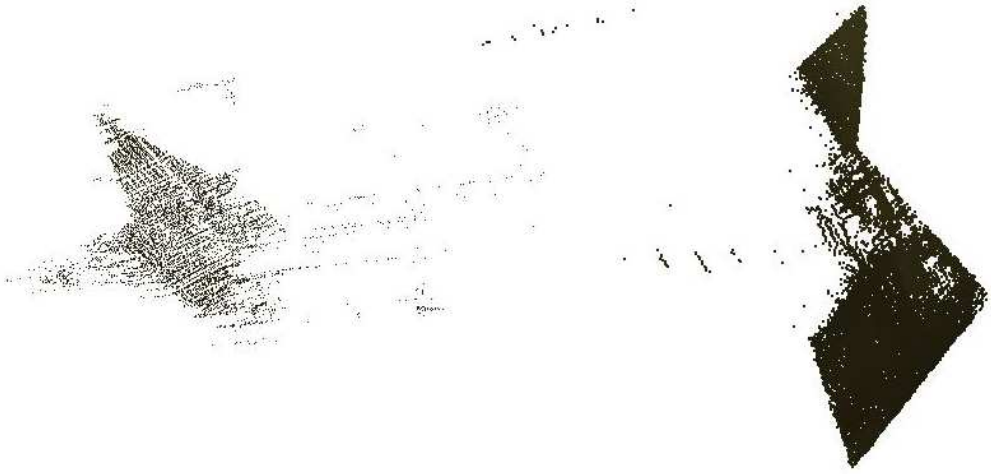
(b) Right side



(c) Left side

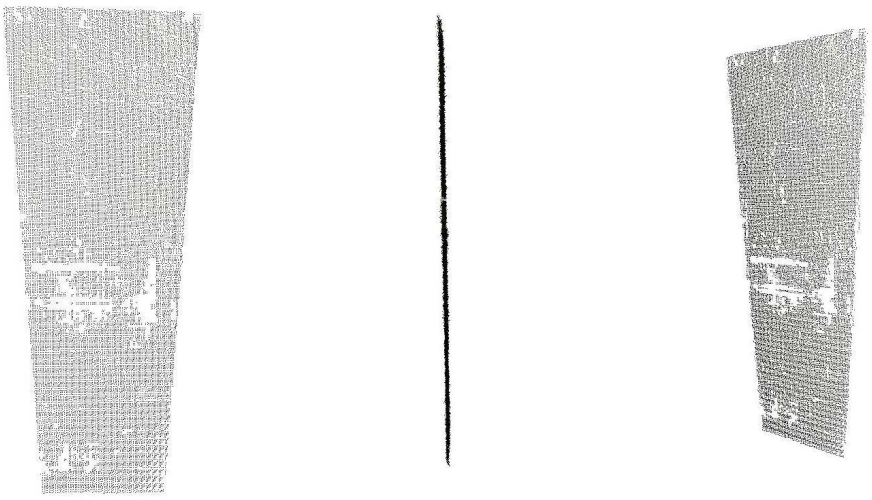
Figure 4.12: Reconstruction results for star trophy using our method, before labelling.

(a) Front

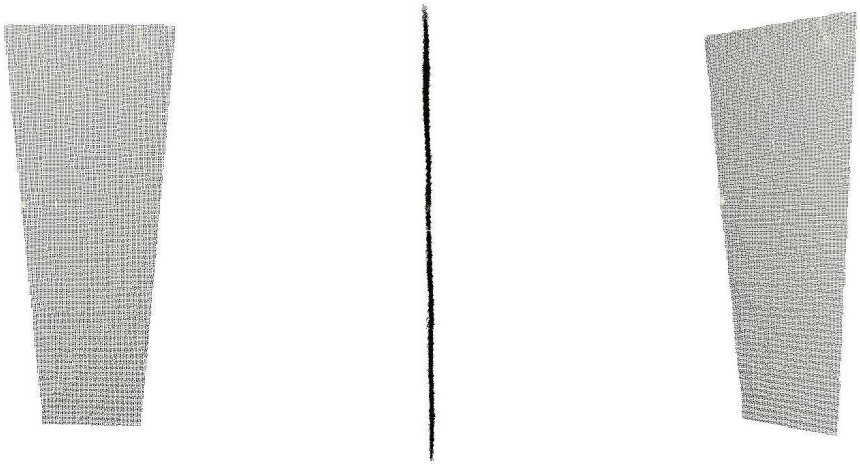


(b) Left side

Figure 4.13: Reconstruction results for star trophy using the Gray code method.



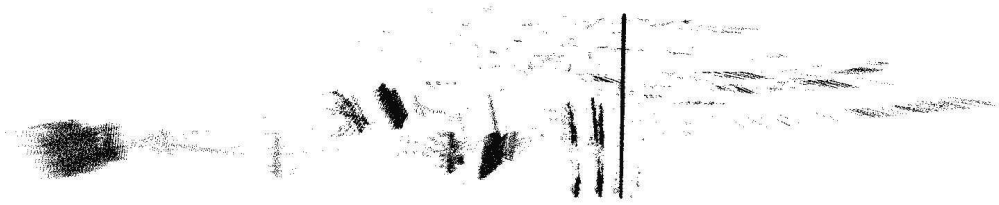
(a) Front (b) Left side (c) Right side



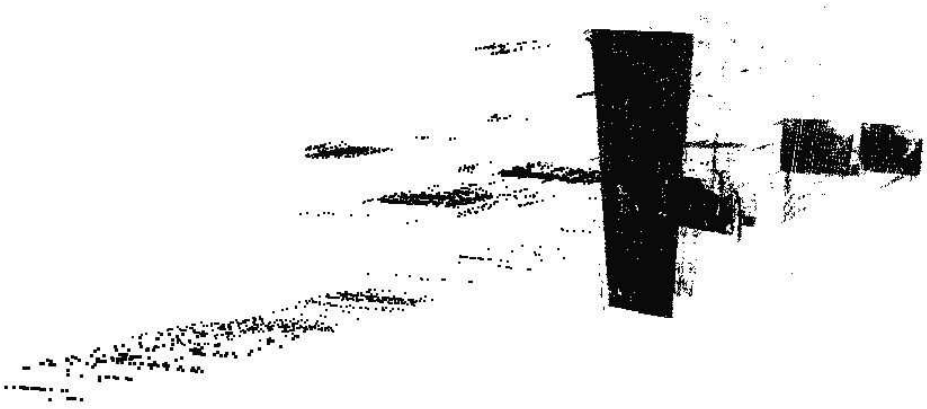
(d) Front (e) Left side (f) Right side

Figure 4.14: Reconstruction results for cone trophy with multiple faces using our method ((a)(b)(c)), compared with the ground truth ((d)(e)(f)).

(a) Front



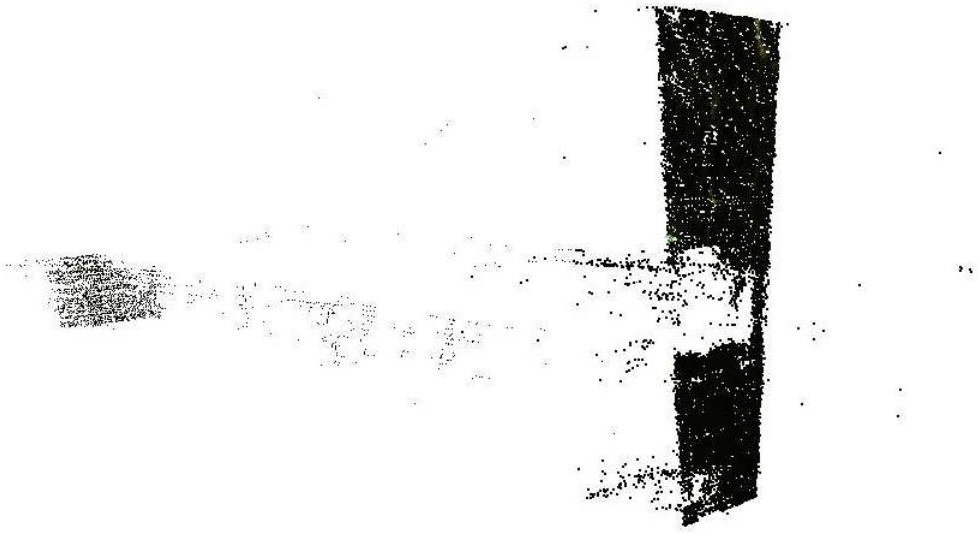
(b) Left side



(c) Right side

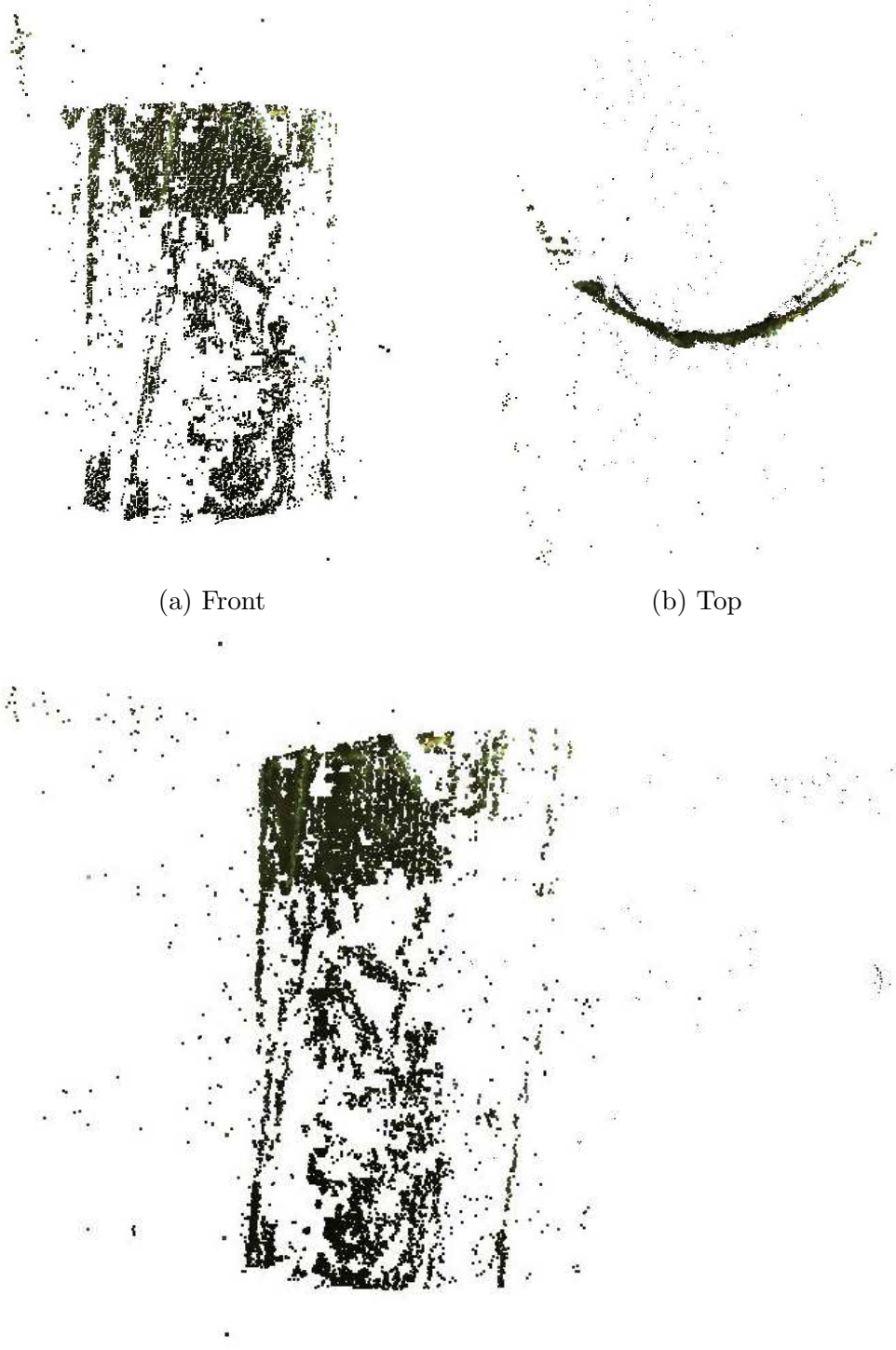
Figure 4.15: Reconstruction results for cone trophy with multiple faces using our method, before labelling.

(a) Front



(b) Left side

Figure 4.16: Reconstruction results for cone trophy with multiple faces using the Gray code method.



(a) Front

(b) Top

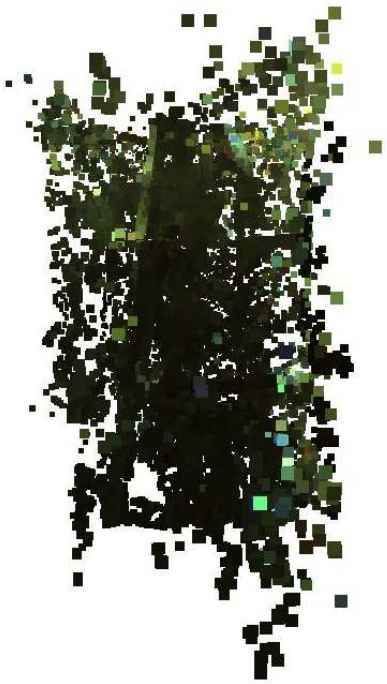
(c) Right side

Figure 4.17: Reconstruction results for big vase using our method.

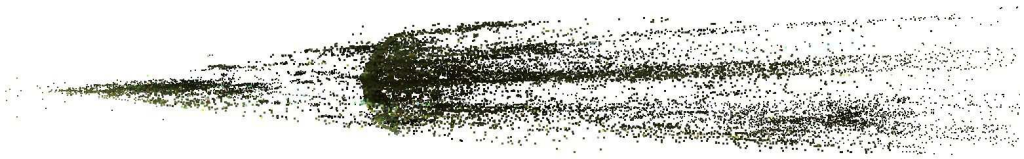
(a) Front

(b) Left side

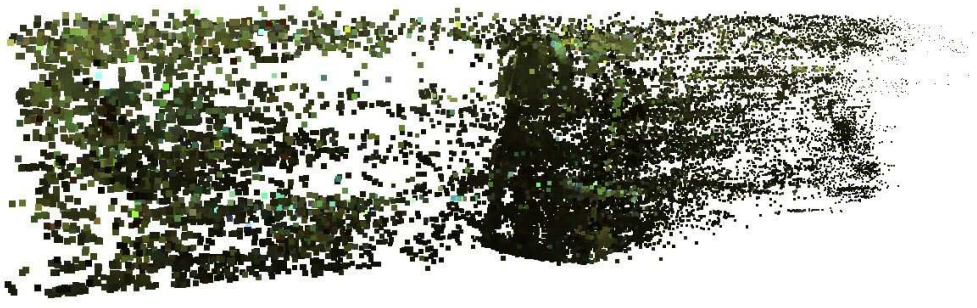
Figure 4.18: Reconstruction results for the big vase using our method, before labelling. The colour denotes the texture of the surface of the object.



(a) Front



(b) Top

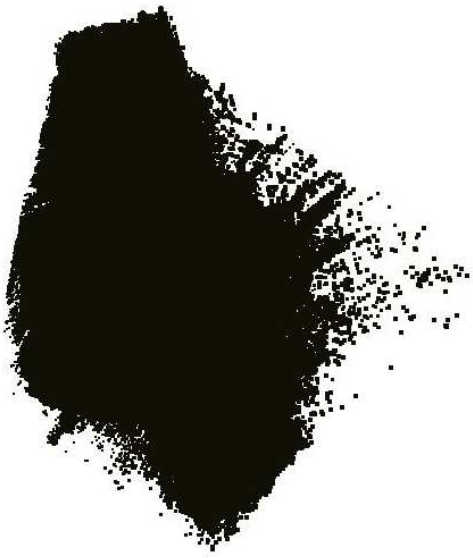


(c) Right side

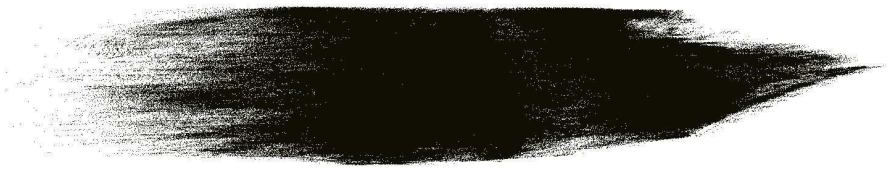
Figure 4.19: Reconstruction results for big vase using the Gray code method.

(a) Front (b) Left side (c) Right side

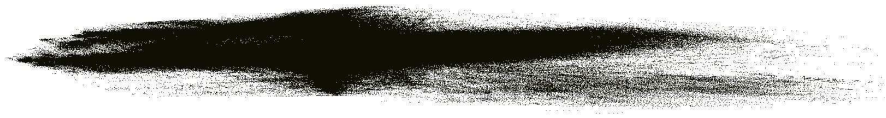
Figure 4.20: Reconstruction results for the small vase using our method.



(a) Front



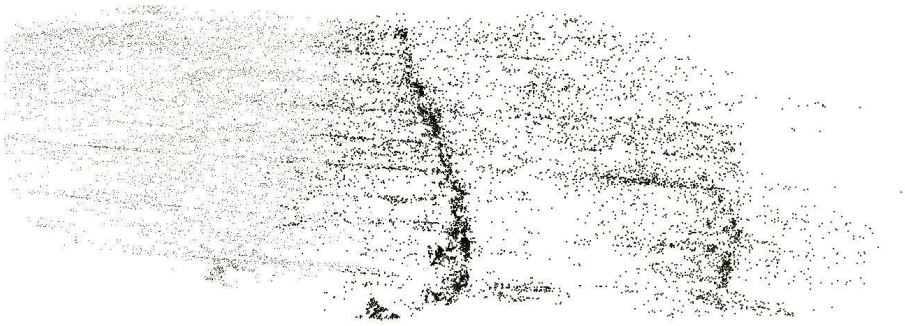
(b) Left side



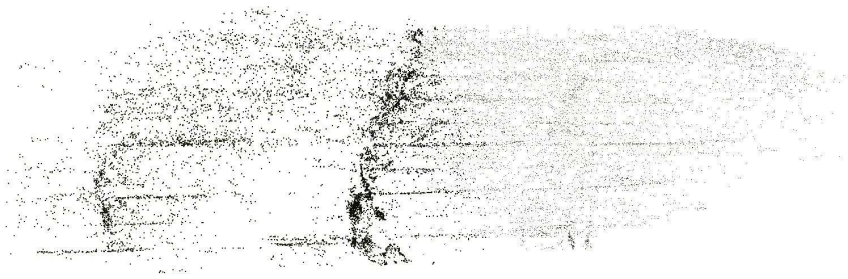
(c) Top

Figure 4.21: Reconstruction results for the small vase using our method, before labelling.

(a) Front

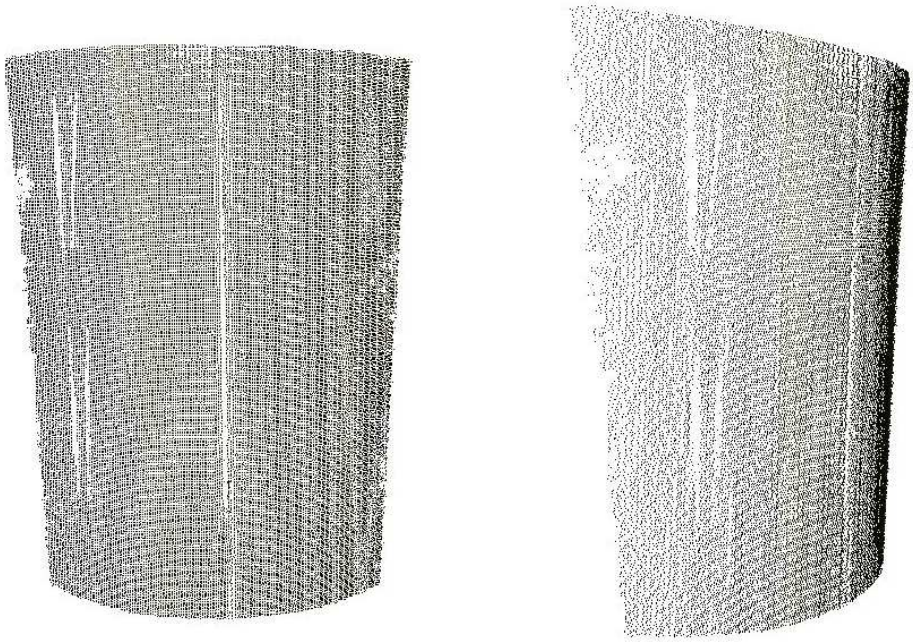


(b) Left side



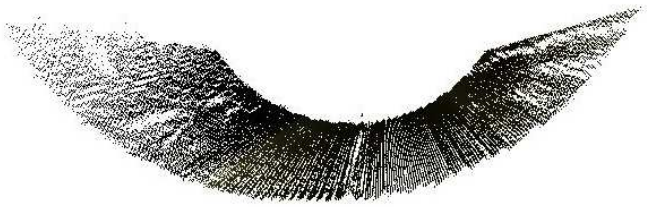
(c) Right side

Figure 4.22: Reconstruction results for the small vase using the Gray code method.



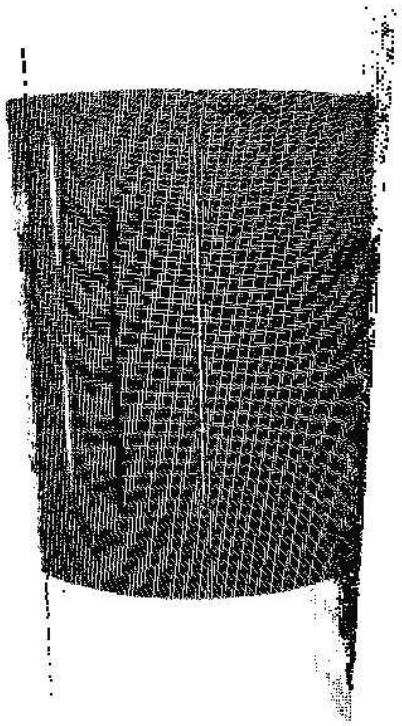
(a) Front

(b) Left side

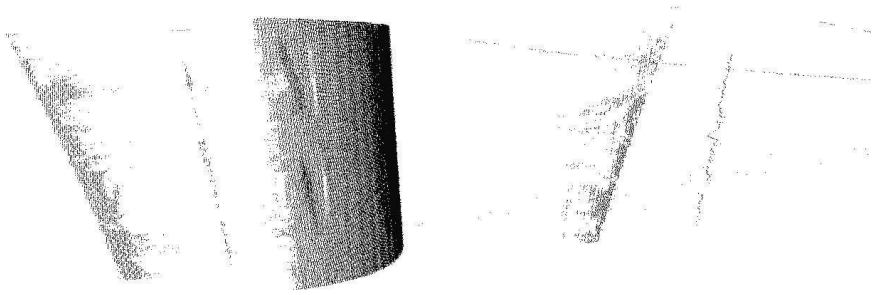


(c) Top

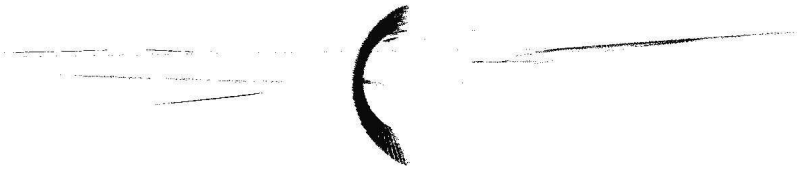
Figure 4.23: Reconstruction results for the anisotropic metal cup using our method.



(a) Front

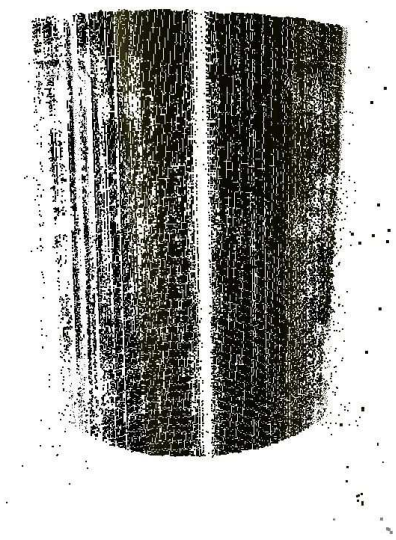


(b) Left side

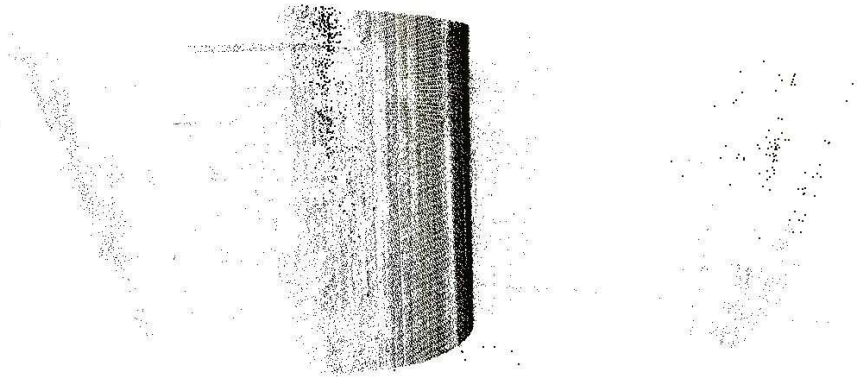


(c) Top

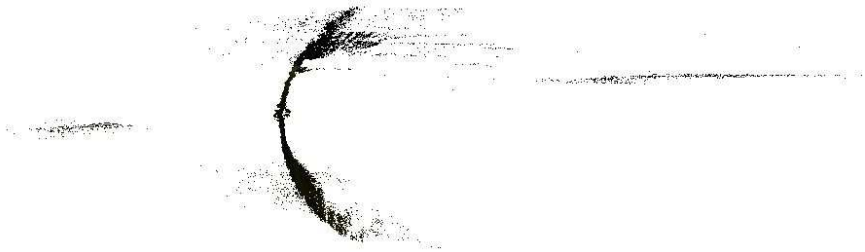
Figure 4.24: Reconstruction results for the anisotropic metal cup using our method, before labelling.



(a) Front

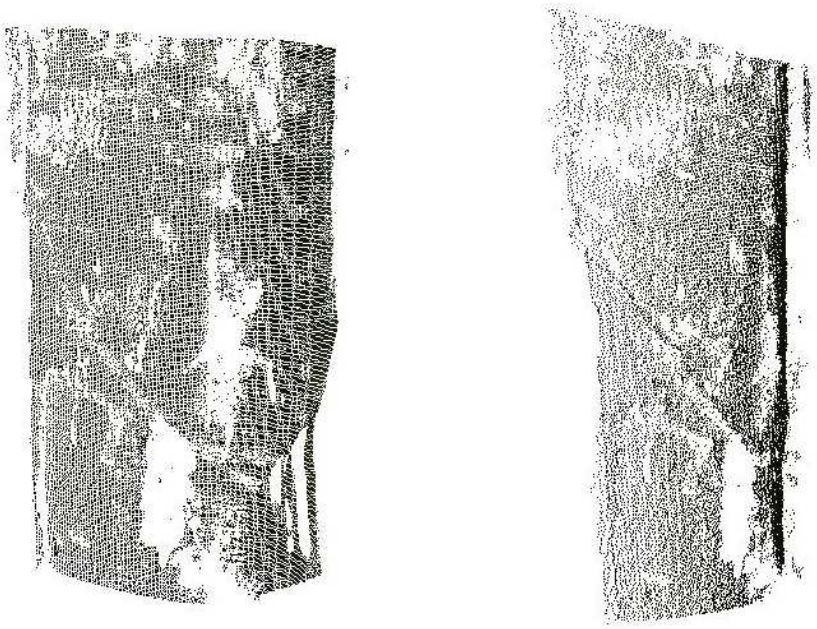


(b) Left side



(c) Top

Figure 4.25: Reconstruction results for the anisotropic metal cup using the Gray code method.

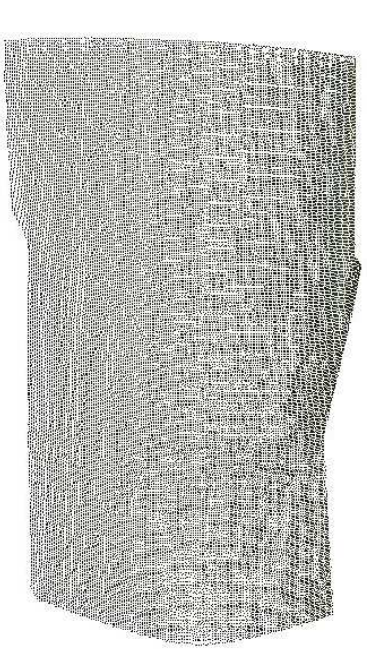


(a) Front

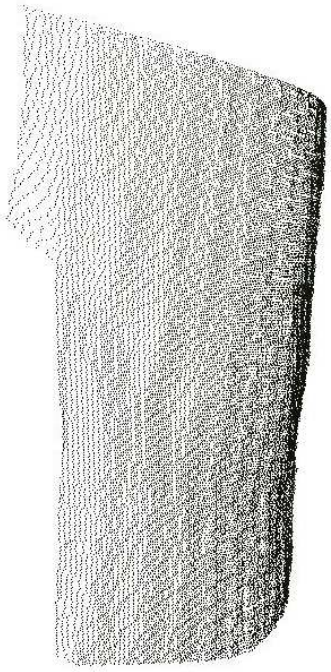
(b) Left side

(c) Top

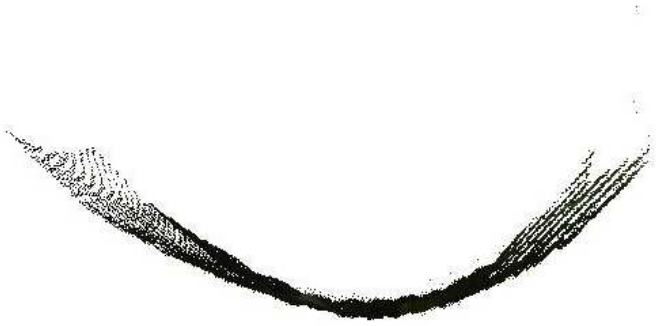
Figure 4.26: Reconstruction results for the plastic cup with two layers using our method.



(a) Front



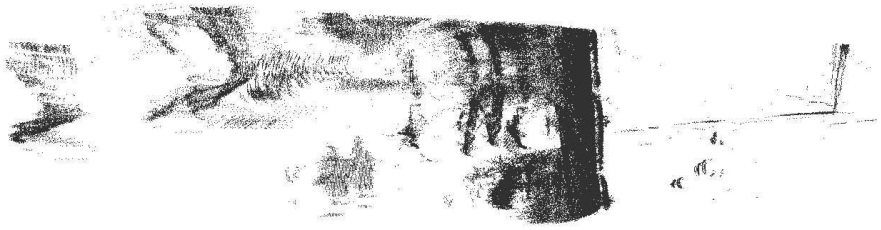
(b) Left side



(c) Top

Figure 4.27: Ground truth for the plastic cup with two layers.

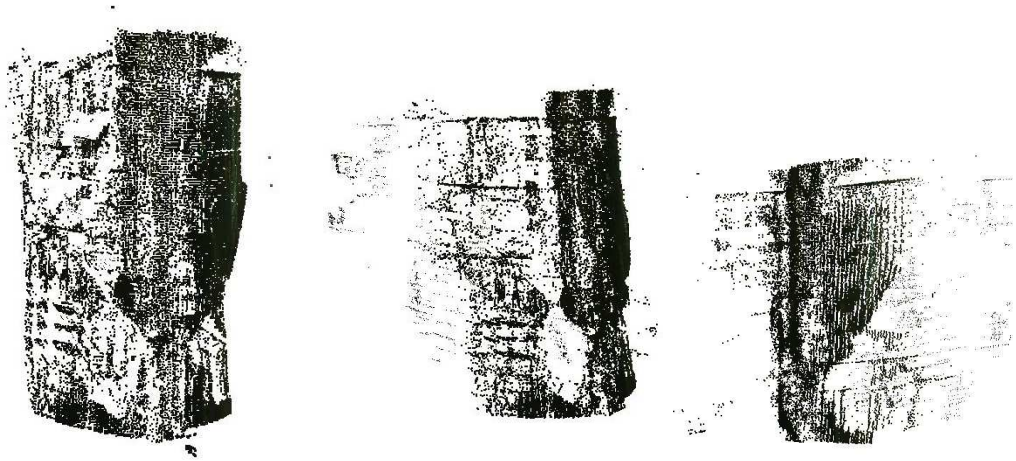
(a) Front



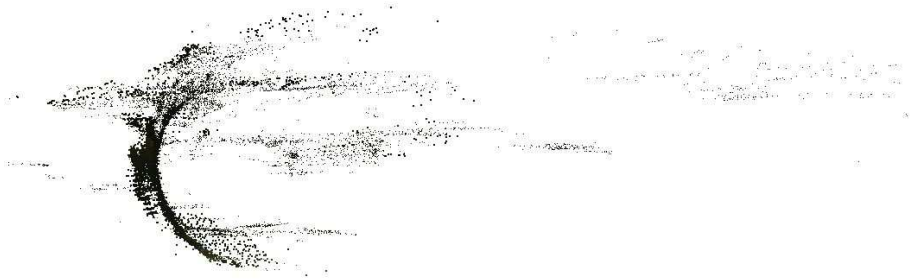
(b) Left side

(c) Top

Figure 4.28: Reconstruction results for the plastic cup with two layers using our method, before labelling.



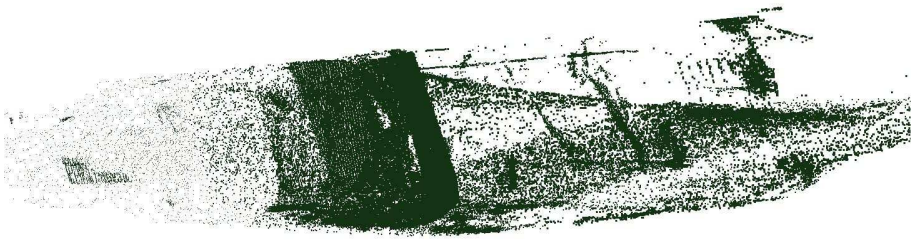
(a) Front (b) Left side (c) Right side



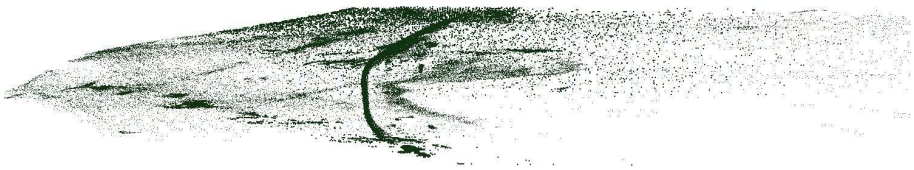
(d) Top

Figure 4.29: Reconstruction results for the plastic cup with two layers using the Gray code method.

(a) Front



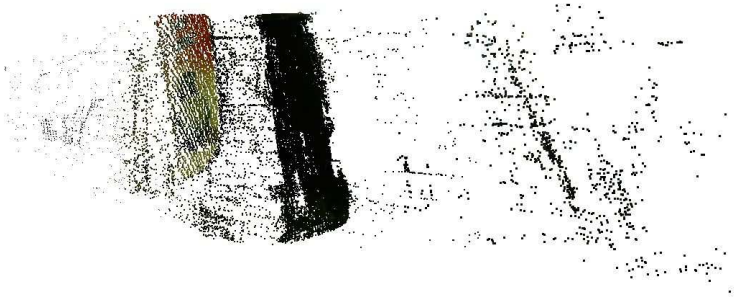
(b) Left side



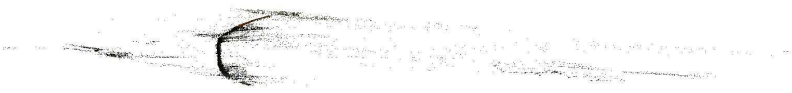
(c) Top

Figure 4.30: Reconstruction results for the plastic bottle with a green dishwashing liquid using our method, before labelling.

(a) Front



(b) Left side



(c) Top

Figure 4.31: Reconstruction results for the plastic bottle with a green dishwashing liquid using the Gray code method.

Chapter 5

Conclusions and Future Work

5.1 Contributions and limitations

In this thesis, a frequency-based method to reconstruct the surface of transparent and specular objects is introduced. Using frequency analysis, complex light composition can be uniquely decomposed without optimization and multiple correspondences between the camera and the projector can be established. Because the new method is based on frequency, which is only determined by the source that creates it and will not be changed by noise, the method is quite robust to noise.

In order to select the correct first-order reflection correspondence from the candidates, a new labelling method is developed. The Markov Random Field is used to define the energy function to be minimized based on the fact that the first-order reflection point is the closest one to the camera center. The Graph Cuts method is used to do the minimization.

A preprocessing technique is used in order to reduce the processing time. An alpha matte is manually extracted by the user to define the region of interest. The processing of the pixels is within this region. A few post processing techniques are also used to improve the final results. Since one point in the patterns corresponds to multiple pixels in the captured images, the average position of these corresponding pixels are used as the corresponding pixel. This strategy makes the reconstructed surface smoother and more accurate.

Some experiments with different objects are conducted and the results are presented and analyzed. For some very challenging objects that previous methods can hardly reconstruct, the new method produces encouraging results. However, for objects with high curvatures or highlighted points, the results of the new method

are not as good as expected.

The main contributions of the proposed frequency-based 3D reconstruction method are summarized as follows:

1. The frequency-based environment matting method is modified for and applied to 3D reconstruction. To our best knowledge, no existing method has ever adapted the frequency-based environment matting method to do 3D reconstruction of transparent and specular objects. The goal of environment matting is to find the correct correspondences between the pixels in the captured images and the points in the displayed patterns, and also the percentages of the contributions, whereas for the 3D reconstruction using structured light, only the correspondences are needed. Hence, methods used to do matting can be adapted to do 3D reconstruction.

2. A labelling method to choose the first-order reflection is introduced and this method is very efficient. Since the transparent and specular objects have a complex interaction with light, multiple correspondences between points in the patterns and pixels in the captured images are established. Hence, when using linear triangulation to find the triangulated points, only one point is on the surface of the object and other points are not. The correct point is called the first-order reflection point. Since the correct point is the closest one to the camera center, when minimizing the distances from the triangulated points to the camera center, the first-order reflection point can be selected. With restrictions such as the closest point and the smoothness of the surface, a labelling method is developed to find the first-order reflection points.

3. The proposed method can not only reconstruct surfaces of transparent objects, but also objects with some anisotropic surfaces. For transparent objects, although it is quite challenging to do 3D reconstruction, given that the objects have a complex interaction with light, with the help of the frequency-based method, candidates of the first-order reflection points can be found. After the labelling procedure, the surface of the object can be reconstructed. For objects with anisotropic surfaces, since the reflection is anisotropic, it is difficult to use the traditional structured light method to do 3D reconstruction. With frequency analysis, it is easy to find the correct correspondences and

straightforward to locate the points on the surface of the objects.

The disadvantages of this new method are listed below.

1. This new method cannot fully deal with highlight. For the point on the surface that receives highlight, the variation of the intensity is very small and the low frequency is difficult to detect because of noise.

2. When the surface of the object has high curvatures, it is difficult to find a good relative position between the camera, the projector and the object. In this case, the camera does not receive enough reflection to find the correct correspondences.

3. For the post processing, an average of the correspondences in the captured images are calculated corresponding to a point in the patterns. This procedure lowers the resolution of the reconstruction result. If the averages are not calculated, one point in the patterns will correspond to multiple pixels in the captured images, which makes the reconstructed surface rough and "thick."

5.2 Recommendations for future work

3D reconstruction for transparent and specular objects is an unsettled and exciting topic in recent years. Because of its wide applications, researchers have devoted many years in order to come up with a more accurate and more efficient method.

The proposed method is based on frequency analysis and is quite robust to noise. However, as discussed before, the problem with highlight cannot be fully solved. One solution is to lower the intensity of the projected patterns to minimize the area of the highlight. This solution may only minimize the highlight, but cannot get rid of it. Another solution is to put the object on a turntable and capture it from different angles. Since the position of the highlight depends on the normal of the point on the surface and the direction of the incoming light, the highlighted point is changing as the turntable rotates. Hence, the highlighted point can be reconstructed when it is not highlighted.

In addition, since the proposed method uses a camera to capture multiple images, it is not quite efficient. Normally, for 1350 images captured at a rate of 1.5 seconds per image, it takes about 33 minutes to capture all the images. One solution is

to use a camcorder as in [4] to largely reduce the acquisition time. A camcorder has a capturing frequency of about 60Hz [86], which means that the camcorder captures 60 images per second. Ideally when the projector can project a video with a frequency of 60Hz, the total image capturing procedure can be only $\frac{1350}{60} = 22.5$ seconds, which is much faster than the proposed method.

For objects with high curvatures, one solution is to rotate the objects for a series of angles and reconstruct the surface one portion at a time. Since the reflection depends on the direction of the incoming light and the normal of the surface, when it comes to surface with high curvatures, the reflected light may be weak when it is received by the camera, making it is difficult to do 3D reconstruction. However, when rotating the object, the reflected light can be received by the camera at an appropriate angle. Using the reflected light at different angles, the surface with high curvatures can be reconstructed.

Another possible future work is to extend the idea of finding the first-order reflections to finding the second-order reflections and so forth. The first-order reflection points consist of the surface of the object. For transparent objects with multiple layers, reconstructing the exterior surface is not enough for the whole structure. The inner structures of the objects are also quite important. Since the inner layers can also reflect the light, the correspondences between the pixels in the captured images and the points in the projected patterns can also be established. After the linear triangulation, the triangulated points are composed of candidates of the first-order reflection points and points on other layers. One basic strategy to find the second and higher order reflections is like peeling an onion. With the help of the labelling procedure, the first-order reflections can be found by the new method. Then, these first-order reflection points are removed from the candidates. After that, the remaining candidates are used as the input to do the labelling procedure again. Since the points in the first layer, i.e. the surface, are removed, the newly calculated points by the labelling procedure actually denote the second layer of the object. This procedure can be done iteratively and ideally multiple layers of the object can be reconstructed step by step. However, this strategy assumes that the reflected light from the inner layers are not refracted before the reflection or after the reflection. When this condition is not met, it is difficult to find the correct correspondences to reconstruct the inner layers.

In the experiments proposed in this thesis, only seven objects are used to do the reconstruction. Future work can include more objects. For the seven objects reconstructed in the experiments, materials such as crystal, plastic, glass and metal have been tried. Structures such as solid object with parallel surfaces, solid objects with multiple faces, objects with complex surface structures, objects with double layers, and objects with inner substances that have different refraction index have also been reconstructed. Although for some of the objects, the reconstruction results are not as good as expected, they are still better than the results of the classic Gray code method. More objects can be included for future experiments, such as the optical disk with anisotropic surface.

Bibliography

[1] Nigel J. W. Morris. *Shape Estimation under General Reflectance and Transparency*. PhD thesis, University of Toronto, 2010.

[2] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, Jun 2011.

[3] Yonatan Wexler, Andrew. W. Fitzgibbon, and Andrew. Zisserman. Image-based environment matting. In *Proceedings of the 13th Eurographics workshop on Rendering*, EGRW '02, pages 279–290, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.

[4] Jiayuan Zhu and Yee-Hong Yang. Frequency-based environment matting. In *Computer Graphics and Applications, 2004. PG 2004. Proceedings. 12th Pacific Conference on*, pages 402–410, 2004.

[5] Ming-June Tsai and Chuan-Cheng Hung. Development of a high-precision surface metrology system using structured light projection. *Measurement*, 38(3):236 – 247, 2005.

[6] J.L Posdamer and M.D Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1 – 17, 1982.

[7] Dalit Caspi, Nahum Kiryati, and Joseph Shamir. Range imaging with adaptive color structured light. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5):470–480, 1998.

[8] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and SrinivasaG. Narasimhan. A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision*, 102(1-3):33–55, 2013.

[9] Fabrice Meriaudeau, L.A. Sanchez Secades, G. Eren, A. Ercil, F. Truchetet, O. Aubreton, and David Fofi. 3-d scanning of nonopaque objects by means of imaging emitted structured infrared patterns. *Instrumentation and Measurement, IEEE Transactions on*, 59(11):2898–2906, 2010.

[10] Matthias B. Hullin, Martin Fuchs, Ivo Ihrke, Hans-Peter Seidel, and Hendrik P. A. Lensch. Fluorescent immersion range scanning. *ACM Transactions on Graphics*, 27(3):87:1–87:10, August 2008.

[11] B. Trifonov, D. Bradley, and W. Heidrich. Tomographic reconstruction of transparent objects. In *Proc. Eurographics Symposium on Rendering*, pages 51–60, 2006.

[12] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Polarization-based transparent surface modeling from two views. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1381–1386. IEEE, 2003.

[13] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Transparent surface modeling from a pair of polarization images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):73–82, 2004.

[14] Sai-Kit Yeung, Tai-Pang Wu, Chi-Keung Tang, T.F. Chan, and S. Osher. Adequate reconstruction of transparent objects on a shoestring budget. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2513–2520, 2011.

[15] Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, Addy Ngan, and Leonard McMillan. Acquisition and rendering of transparent and refractive objects. In *Proceedings of the 13th Eurographics workshop on Rendering*, EGRW '02, pages 267–278, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.

[16] Yuanyuan Ding, Feng Li, Yu Ji, and Jingyi Yu. Dynamic fluid surface acquisition using a camera array. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2478–2485, 2011.

[17] N.J.W. Morris and K.N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 2007.

[18] R.J. Valkenburg and A.M. McIvor. Accurate 3d measurement using a structured light system. *Image and Vision Computing*, 16:99–110, 1996.

[19] Ivo Ihrke, Kiriakos N Kutulakos, Hendrik PA Lensch, Marcus Magnor, and Wolfgang Heidrich. State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS 2008 STAR–STATE OF THE ART REPORT*. Citeseer, 2008.

[20] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[21] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[22] M. Brown and D.G. Lowe. Unsupervised 3d object recognition and reconstruction in unordered datasets. In *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, pages 56–63, 2005.

[23] Keju Peng, Xin Chen, Dongxiang Zhou, and Yunhui Liu. 3d reconstruction based on sift and harris feature points. In *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on*, pages 960–964, 2009.

[24] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359, 2008. ¡ce:title¿Similarity Matching in Computer Vision and Multimedia¡/ce:title¿.

[25] Jianguo Li, E. Li, Yurong Chen, Lin Xu, and Yimin Zhang. Bundled depth-map merging for multi-view stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2769–2776, 2010.

[26] Chris Hermans, Yannick Francken, Tom Cuypers, and Philippe Bekaert. Depth from sliding projections. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1865–1872. IEEE, 2009.

[27] JL Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17, 1982.

[28] Harold Fredricksen. The lexicographically least de bruijn cycle. *Journal of Combinatorial Theory*, 9(1):1–5, 1970.

[29] Jordi Pagès, Joaquim Salvi, Christophe Collewet, and Josep Forest. Optimised de bruijn patterns for one-shot shape acquisition. *Image and Vision Computing*, 23(8):707–720, 2005.

[30] Philipp Fechteler and Peter Eisert. Adaptive colour classification for structured light systems. *Computer Vision, IET*, 3(2):49–59, 2009.

[31] HB Wu, Y Chen, MY Wu, CR Guan, and XY Yu. 3d measurement technology by structured light using stripe-edge-based gray code. In *Journal of Physics: Conference Series*, volume 48, page 537. IOP Publishing, 2006.

[32] Daniel G Aliaga and Yi Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[33] M. Gupta, A. Agrawal, A. Veeraraghavan, and S.G. Narasimhan. Structured light 3d scanning in the presence of global illumination. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 713–720, 2011.

[34] Peisen S. Huang and Song Zhang. Fast three-step phase-shifting algorithm. *Appl. Opt.*, 45(21):5086–5091, Jul 2006.

[35] Peisen S. Huang, Song Zhang, and Fu-Pen Chiang. Trapezoidal phase-shifting method for three-dimensional shape measurement. *Optical Engineering*, 44(12):123601–123601–8, 2005.

[36] Song Zhang. Recent progresses on real-time 3d shape measurement using digital fringe projection techniques. *Optics and Lasers in Engineering*, 48(2):149 – 158, 2010.

[37] Song Zhang and Peisen Huang. High-resolution, real-time 3d shape acquisition. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pages 28–28. IEEE, 2004.

[38] Miguel Arevallilo Herráez, David R Burton, Michael J Lalor, and Munther A Gdeisat. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Applied Optics*, 41(35):7437–7444, 2002.

[39] Antonio Baldi. Two-dimensional phase unwrapping by quad-tree decomposition. *Appl. Opt.*, 40(8):1187–1194, Mar 2001.

[40] Miguel Arevallilo Herráez, Munther A Gdeisat, David R Burton, and Michael J Lalor. Robust, fast, and effective two-dimensional automatic phase unwrapping algorithm based on image decomposition. *Applied Optics*, 41(35):7445–7455, 2002.

[41] Song Zhang. Composite phase-shifting algorithm for absolute phase measurement. *Optics and Lasers in Engineering*, 50(11):1538–1541, 2012.

[42] Giovanna Sansoni, Sara Lazzari, Stefano Peli, and Franco Docchio. 3d imager for dimensional gauging of industrial workpieces: state of the art of the development of a robust and versatile system. In *3-D Digital Imaging and Modeling, 1997. Proceedings., International Conference on Recent Advances in*, pages 19–26. IEEE, 1997.

[43] Georg Wiora. High-resolution measurement of phase-shift amplitude and numeric object phase calculation. In *International Symposium on Optical Science and Technology*, pages 289–299. International Society for Optics and Photonics, 2000.

[44] Jens Gühring. Dense 3d surface acquisition by structured light using off-the-shelf components. In *Photonics West 2001-Electronic Imaging*, pages 220–231. International Society for Optics and Photonics, 2000.

[45] Nigel J. W. Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. In *Proceedings of the Tenth IEEE International Conference on Computer Vision*

- *Volume 2*, ICCV '05, pages 1573–1580, Washington, DC, USA, 2005. IEEE Computer Society.

[46] Thorsten Bothe, Wansong Li, Christoph von Kopylow, and Werner P. O. Juptner. High-resolution 3d shape measurement on specular surfaces by fringe reflection. pages 411–422, 2004.

[47] Kiriakos N. Kutulakos and Eron Steger. A theory of refractive and specular 3d shape by light-path triangulation. *Int. J. Comput. Vision*, 76(1):13–29, January 2008.

[48] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive shape from light field distortion. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1180–1186, 2011.

[49] Sameer Agarwal. Refractive height fields from single and multiple images. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 286–293, Washington, DC, USA, 2012. IEEE Computer Society.

[50] Yung-Yu Chuang, Douglas E. Zongker, Joel Hindorff, Brian Curless, David H. Salesin, and Richard Szeliski. Environment matting extensions: Towards higher accuracy and real-time capture. In *Proceedings of ACM SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, pages 121–130. ACM Press / ACM SIGGRAPH / Addison Wesley Logman, July 2000. ISBN 1-58113-208-5.

[51] Yoav Y Schechner, Shree K Nayar, and Peter N Belhumeur. A theory of multiplexed illumination. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 808–815. IEEE, 2003.

[52] Kazutada Koshikawa. A polarimetric approach to shape understanding of glossy objects. In *Proceedings of the 6th international joint conference on Artificial intelligence - Volume 1*, IJCAI'79, pages 493–495, San Francisco, CA, USA, 1979. Morgan Kaufmann Publishers Inc.

[53] Daisuke Miyazaki and Katsushi Ikeuchi. Shape estimation of transparent objects by using inverse polarization ray tracing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(11):2018–2030, 2007.

[54] Diego Nehab, Tim Weyrich, and Szymon Rusinkiewicz. Dense 3d reconstruction from specularity consistency. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[55] Yannick Francken, Tom Cuypers, Tom Mertens, Jo Gielis, and Philippe Bekaert. High quality mesostructure acquisition using specularities. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7. IEEE, 2008.

[56] Yannick Francken, Tom Cuypers, and Philippe Bekaert. Mesostructure from specularity using gradient illumination. In *Proceedings of the 5th ACM/IEEE International Workshop on Projector camera systems*, page 11. ACM, 2008.

[57] Yannick Francken, Tom Cuypers, Tom Mertens, and Philippe Bekaert. Gloss and normal map acquisition of mesostructures using gray codes. In *Advances in Visual Computing*, pages 788–798. Springer, 2009.

[58] Thomas Porter and Tom Duff. Compositing digital images. *SIGGRAPH Comput. Graph.*, 18(3):253–259, January 1984.

[59] Yung-Yu Chuang, B. Curless, D.H. Salesin, and R. Szeliski. A bayesian approach to digital matting. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–264–II–271 vol.2, 2001.

[60] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):228–242, February 2008.

[61] Jian Sun, Jiaya Jia, Chi-Keung Tang, and Heung-Yeung Shum. Poisson matting. *ACM Trans. Graph.*, 23(3):315–321, August 2004.

[62] A. Levin, A. Rav Acha, and D. Lischinski. Spectral matting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1699–1712, 2008.

[63] Douglas E. Zongker, Dawn M. Werner, Brian Curless, and David H. Salesin. Environment matting and compositing. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 205–214, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[64] Q. Duan, Jianfei Cai, and Jianmin Zheng. Vector field fitting for real-time environment matting of transparent objects. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3229–3232, 2009.

[65] Q. Duan, Jianfei Cai, Jianmin Zheng, and Weisi Lin. Fast environment matting extraction using compressive sensing. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6, 2011.

[66] Sai-Kit Yeung, Chi-Keung Tang, Michael S. Brown, and Sing Bing Kang. Matting and compositing of transparent and refractive objects. *ACM Trans. Graph.*, 30(1):2:1–2:13, February 2011.

[67] Pieter Peers and Philip Dutré. Wavelet environment matting. In *Proceedings of the 14th Eurographics workshop on Rendering*, EGRW '03, pages 157–166, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[68] Biswarup Choudhury, Deepali Singla, and Sharat Chandran. Fast color-space decomposition based environment matting. In *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, I3D '08, pages 1:1–1:1, New York, NY, USA, 2008. ACM.

[69] Oliver Goretzki. 3d reconstruction of medical images using java3d. *Bachelor Thesis, Luxembourg.*

[70] Nina Jährling, Klaus Becker, Cornelia Schönbauer, Frank Schnorrer, and Hans-Ulrich Dodt. Three-dimensional reconstruction and segmentation of intact drosophila by ultramicroscopy. *Frontiers in systems neuroscience*, 4, 2010.

[71] Olivier Morel, Christophe Stolz, Fabrice Meriaudeau, and Patrick Gorria. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Appl. Opt.*, 45(17):4062–4068, Jun 2006.

[72] Olivier Morel, Fabrice Meriaudeau, Christophe Stolz, and Patrick Gorria. Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In *Electronic Imaging 2005*, pages 178–186. International Society for Optics and Photonics, 2005.

[73] Juan A Barceló. Virtual reality for archaelogical explanation. beyond" picturesque" reconstruction. *Archeologia e Calcolatori*, (12):221–244, 2001.

[74] Fabio Bruno, Stefano Bruno, Giovanna De Sensi, Maria-Laura Luchi, Stefania Mancuso, and Maurizio Muzzupappa. From 3d reconstruction to virtual reality: A complete methodology for digital archaeological exhibition. *Journal of Cultural Heritage*, 11(1):42–49, 2010.

[75] Wikipedia. Discrete fourier transform — Wikipedia, the free encyclopedia, 2013. [Online; accessed June-2013].

[76] James W. Cooley, Peter A W Lewis, and Peter D. Welch. The fast fourier transform and its applications. *Education, IEEE Transactions on*, 12(1):27–34, 1969.

[77] THE GIMP TEAM. Gimp 2.8, 2013. [Online; accessed June-2013].

[78] Richard I Hartley and Peter Sturm. Triangulation. *Computer vision and image understanding*, 68(2):146–157, 1997.

[79] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2000.

[80] Xida Chen, Yufeng Shen, and Yee Hong Yang. Background estimation using graph cuts and inpainting. In *Proceedings of Graphics Interface 2010*, GI '10, pages 97–103, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.

[81] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6):1068–1080, 2008.

[82] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[83] MeshLab. *version 1.3.2*. The MeshLab Team, 2012.

[84] Gabriel Falcao, Natalia Hurtos, and Joan Massich. Plane-based calibration of a projector-camera system. *VIBOT Master*, 9, 2008.

[85] Jean-Yves Bouguet. Camera calibration toolbox for matlab. 2004.

[86] Wikipedia. List of panasonic camcorders — Wikipedia, the free encyclopedia, 2013. [Online; accessed 11-July-2013].