

Frequency-Based 3D Reconstruction of Transparent and Specular Objects

Ding Liu Xida Chen Yee-Hong Yang

Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

{dliu5, xida, herberty}@ualberta.ca

Abstract

3D reconstruction of transparent and specular objects is a very challenging topic in computer vision. 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 either passive stereo or the traditional structured light methods to do the reconstruction. We propose a frequency-based 3D reconstruction method, which incorporates the frequency-based matting method. Similar to the structured light methods, a set of frequency-based patterns are projected onto the object, and a camera captures the scene. Each pixel of the captured image is analyzed along the time axis and the corresponding signal is transformed to the frequency-domain using the Discrete Fourier Transform. 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, the surface of transparent and specular objects can be reconstructed with very encouraging results.

1. Introduction

3D reconstruction is the procedure of capturing the shape or surface structure of an object. The goal is to acquire the 3D information of each point on the surface of an object. For opaque objects with Lambertian surface, many methods [8, 21] can be used. However, for objects with poor reflection or anisotropic properties (Fig. 1a and 1b), which means that the reflection is either weak or non-uniform, 3D reconstruction is still an active research topic.

Traditionally, methods for 3D reconstruction of opaque objects use structured light with coded patterns, but these methods may fail for transparent and specular objects. For transparent objects, the projected patterns will transmit through the object and reflected by the background, which may interfere with the reflection from the object surface.

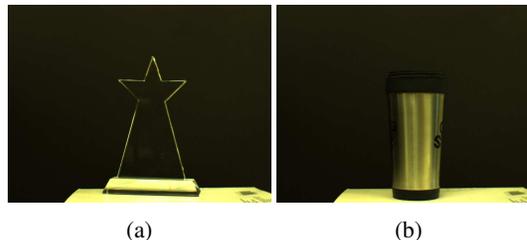


Figure 1: Examples of objects that are difficult to be reconstructed using existing methods. (a) A transparent trophy with weak reflection. (b) A metal cup with anisotropic properties.

For specular objects, the reflection is view-dependent and sometimes the intensity of the reflection is very strong, which may also interfere with the projected pattern. The interference makes it very difficult to find the correct correspondences between points on the pattern and pixels on the images. However, the frequency-based environment matting [24] can be adapted to accurately find the correct correspondences. To our best knowledge, none of the existing methods have incorporated this method into 3D reconstruction for the purpose of finding the correct correspondences.

The goal of our work is to use the structured light methods incorporated with the environment matting method to perform 3D reconstruction of transparent and specular objects. Based on the challenges stated above, our contributions are as follows. First, the proposed method incorporates the environment matting method for 3D reconstruction, and can find the correct correspondences between points on the projected patterns and pixels on the 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 objects and specular objects with anisotropic surface.

2. Related Work

3D reconstruction was originally introduced to acquire the shape and surface structure of diffuse opaque objects. One of the fundamental methods is the structured light

method using code patterns such as gray code and phase shifting [7]. The advantage of these method is the resilience to imaging errors and noise as well as the simple setup.

Despite its high accuracy, the structured light methods are seldom directly applied to the 3D reconstruction of transparent and specular objects. The main reason is the active optical interaction of the objects with light. Researchers have been working in this area for over three decades and many effective methods have been proposed. These methods can be categorized into several groups. The first group of methods recover the shape from distortion. The early work [16, 17] mainly reconstructs the surface of waving water and recovers the pattern under the water. The results are highly dependent on the accuracy of the optical flow and, to acquire accurate results, the observation time can be very long. A checkerboard pattern is usually used and observed after distorted by the transparent objects [5, 14]. The correspondences between the pattern and the captured image are estimated, and the 3D information of surface points is obtained after an optimization procedure. Other than a checkerboard pattern, a light field probe can also be used [11, 23]. Ji *et al.* [11] observe the light field probe through a gas flow, acquire a dense set of ray-ray correspondences and then reconstruct their light paths. These methods are limited to a single refractive event along each camera ray [23] to simplify the estimation procedure. For specular objects, shape from distortion can be applied as well [18]. The method is based on the consistency of surface normal and depth. By minimizing a stereo matching cost function, the surface normal and depth can be estimated. However, because of the limitation of the setup, the method can only recover surface normal within a small range of angles, which is limited to objects with relatively flat and smooth surface. The methods in the second category use direct ray measurement. These methods are based on the measurement of calibrated planar targets imaged at different positions with respect to the object to be reconstructed [12] or approximated from optical flow data [2]. Although these methods can perform the reconstruction of dynamic effects, they need more than one camera to capture the scene simultaneously. Most importantly, they cannot handle cases when the object has a complex interaction with light. The third category includes the reflectance-based reconstruction methods. A typical method uses scatter traces to recover the surface of inhomogeneous transparent objects [15]. This method needs to move the setup during the experiment. Since it is very sensitive to the calibration of the light source and the camera, the movement of the light source will introduce errors to the results. Hence, the results are often difficult, if not impossible, to reproduce. To our best knowledge, no other researchers, including ourselves, were able to obtain satisfactory results. The fourth category takes polarization into account [13]. These methods use the property that the

object's surface normal can be determined by the degree of polarization of the light reflected from the object's surface. However, these methods can only be applied well to a small group of objects that have a simple shape with known refractive index. The fifth group uses a tomographic method which usually requires the object to be suspended in a highly toxic solution [20]. By matching the refractive index of the solution with that of the transparent object, the refractive effect can be minimized. A major issue is that the solution is toxic and requires extreme caution during experiments. Also, it may damage the object. The methods in the last category uses direct sampling by altering the immersing medium. Both [10, 6] can be applied to reconstruct challenging transparent objects with complex surface structures. However, the major limitation is their invasive process, which requires physical contact between the object and the medium. For example, the method [10] immerses the object into a fluorescent liquid, which may cause chemical reaction with the object or even damage it. The method [6] uses laser to heat up the surface and reconstructs the surface by thermal images. The heating procedure alone can be time-consuming and the heat may damage the object as well. Specifically for specular objects, shape from specular-ity is introduced to recover the texture and fine-scale shape [22]. The method uses a BRDF/BTF measurement device to recover surface normal as well as spatially varying BRDFs. The limitation is that its setup is very complicated, which requires camera, light source with moving aperture, collimating lens assembly, beam splitter, and off-axis concave parabolic mirror.

The method presented in [1] is probably the most related to our method. In particular, the method combines the binary and frequency based structured light patterns to estimate the pixel correspondences. The method consists of two steps. In the first step, it partitions the projected pattern into small rectangular tiles and establishes a corresponding tile for each pixel in the image. An inter-tile binary coding scheme is presented to assign a unique code to each tile. After that, an intra-tile coding strategy is used to resolve the ambiguity inside each tile. Because of the their coding techniques, the number of captured images is less than our method which uses the frequency-based coding [24]. However, the method extract all the possible correspondences for each pixel only and there is no further analysis of these correspondences. Comparing to that, our method analyzes the potential correspondences so that multiple layers of complex objects can be extracted.

As discussed above, the traditional structured light method requires a very simple setup but cannot be directly applied to the 3D reconstruction of transparent and specular objects. The methods proposed specifically for transparent objects have certain limitations. Therefore, our goal is to combine the advantages of both. In particular, we

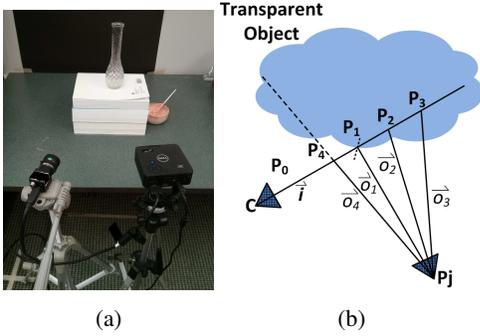


Figure 2: (a) Experimental setup. (b) An illustration of multiple intersections.

present a new method to reconstruct the surface of transparent objects with the structured light setup. The main challenge for 3D reconstruction of transparent objects is to find the correct correspondences for triangulation. The goal of environment matting is to find the correct correspondences between the displayed pattern and the captured image. Zhu and Yang [24] introduce an elegant method called the frequency-based environment matting method, which was inspired by the fact that a time domain signal has a unique decomposition in the frequency domain. After transforming the captured signals into the frequency domain, unique correspondences between the backdrop patterns and the obtained images can be established. Our 3D reconstruction method incorporates this method and achieves very encouraging results.

3. Proposed Method

3.1. Setup

To clearly explain our method, the setup is discussed first. Shown in Fig. 2a, our setup is similar to the traditional structured light setup. The projector and the camera are located on the same side of the object. The object is put in front of a black cloth to minimize the interference of 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. Environment matting

The first step of the proposed method is modified from the frequency-based environment matting method [24].

3.2.1 Pattern Design

The projected patterns are designed as follows

$$I(i, t) = [\cos(2\pi \cdot (i + 10) \cdot t) + 1] \cdot 120 \quad (1)$$

where each pixel position on the patterns has an intensity $I(i, t)$, which varies with 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. Interested readers should refer to [24] for details of generating patterns.

3.2.2 Analyzing the captured images

In order to accelerate the processing time, we manually extract the region of interest. Since the scene is fixed, our method is applied only to the region of interest.

The images are stacked and analyzed in the frequency domain. The key property of frequency, which is also the reason for using the frequency analysis, is that the frequency of a signal only relies on the source that creates it, and is not changed by the medium. According to [24], frequency analysis has two additional desirable properties: ❶ Different frequencies of a signal will show up in the frequency domain. ❷ It is robust to noise. More details are described in [24] for analyzing the captured images.

We apply 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 all these peaks have most of the contributions to the converged point and so they are all selected as candidates of the “first-order reflections,” which correspond to the points on the front surface of the object. The reason of not choosing the global maximum as the first-order reflection is because the object is transparent, and a major portion of light is transmitted into the object. Only a small portion is reflected directly from the surface. Hence, the first-order reflected light may not contain the most energy, and so in the corresponding power spectrum, it may not be the maximum globally. However, comparing to the power spectra of other neighbouring pixels, the first-order reflection should be locally maximum.

After finding the local maxima of the power spectrum, their corresponding frequencies can be determined. These frequencies uniquely locate a group of potential correspondences on the projected pattern for each image pixel. Using linear triangulation, a set of 3D points can be computed from all the potential correspondences. These 3D points are candidates of points on the surface of the object, but only one of them is correct, which is the desired first-order reflection point. We select the first-order reflection point from these candidates using a new labelling procedure.

3.3. Labelling

Fig. 2b 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 P_j with the con-

tributing pixels in the pattern can determine multiple directions, \vec{o}_1 , \vec{o}_2 , \vec{o}_3 and \vec{o}_4 . These directions intersect the incoming direction at P_1 , P_2 , P_3 , and P_4 .

Among these intersections, intuitively the one closest to the camera center should be the first-order reflection point. However, there is an exception. Although P_4 is closer to the camera than P_1 , it is not the first-order reflection point. Because as shown in Fig. 2b, the direction \vec{o}_4 first refracts into the object, and after several 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 may be so small that it may not satisfy the local maximum selection criterion. Hence, P_4 is not likely chosen as one of the candidates.

Now the conclusion is that the first-order reflection point should be the closest 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.* [4] to do labelling.

3.3.1 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 minimizes the total energy. Similar to [4], the energy function is defined based on the Markov Random Field as

$$E(f_p) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{p,q}(f_p, f_q), \quad (2)$$

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

3.3.2 Data term

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

$$D_p(f_p) = \sqrt{\sum_{i=x,y,z} (f_{pi} - C_i)^2}, \quad (3)$$

where f_{pi} denotes the i th value of the 3D coordinates of the pixel p , after assigning the label f_p to it. C_i is the i th value of the 3D coordinates of the camera center, which is chosen as the origin of the coordinate system.

3.3.3 Smoothness term

In Eq. 2, $V_{p,q}(f_p, f_q)$ 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 smoothness term of a pixel with each of its neighbor is calculated using

$$V_{p,q}(f_p, f_q) = |D_p(f_p) - D_q(f_q)|, \quad (4)$$

where $D_p(f_p)$ is the data term of label f_p . $|\cdot|$ denotes the absolute value, indicating the difference of the distances from the camera center to the two triangulated points. The pixel q is one of the neighbors of the pixel p in the captured image. In this paper, each pixel is assumed to have 8 neighbors.

3.3.4 Minimizing the energy function

The energy function is minimized using the classical graph cuts method. According to the results from [19], the expansion move algorithm introduced by [3] gives faster and better results than other methods in general. Hence, it is chosen to minimize the energy function.

3.4. Post-processing

After labelling, the first-order reflection points can be determined 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.’’ Hence, for each pixel in the pattern, we find which pixels in the camera are mapped to it and use the average position as the correct correspondence of the captured image.

4. Experimental Results

In our experiments, seven objects are used in the reconstruction. 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 refractive indices have also been reconstructed. The classical gray code method is used to do 3D reconstruction and the results are accordingly compared with those using the proposed method. 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

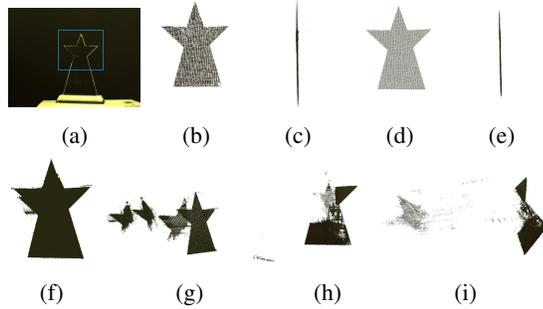


Figure 3: Star trophy. (a) The object. (b)(c) The reconstruction result using our method, seen from the front and left side. (d)(e) The ground truth, seen from the front and left. (f)(g) The result before labelling, seen from the front and left. (h)(i) The result using the gray code method, seen from the front and left.

gently brushed onto the object. After the paint is dried, the gray code method is used to reconstruct the “opaque” object and the result is used as the ground truth. 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.

4.1. 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 a green dishwashing liquid in it.

For the star trophy (Fig. 3) and the cone trophy with multiple faces (Fig. 4), the objects are solid and 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 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 first-order reflection candidates. The traditional methods using structured light fail because of the complex optical interactions, shown in Fig. 3h, 3i, 4h and 4i. However, using the proposed method, good results can be acquired (Fig. 3 and 4). The surface of the objects is reconstructed smoothly, although there are a few small holes in the results, because of the highlight. For pixels with strong highlight, their intensities have little variations. Hence, when transformed to the frequency domain, the magnitude of corresponding frequency can be as low as noise. Hence, the pixels in the highlight region may be wrong or no correspondence may be obtained.

Fig. 5 shows the reconstruction results of a small vase that has detailed “pineapple” textures on the surface. Strong

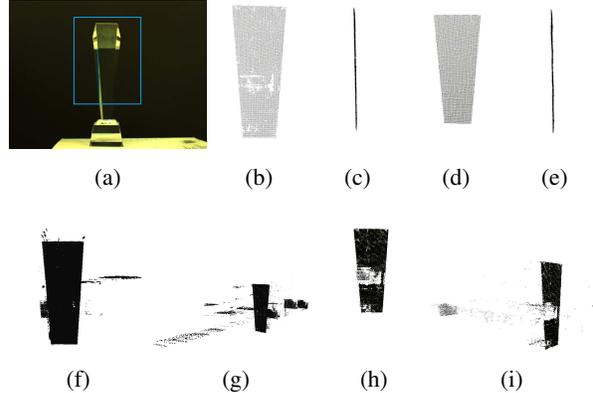


Figure 4: Cone trophy with multiple faces. (a) The object. (b)(c) The reconstruction result using our method, seen from the front and left side. (d)(e) The ground truth, seen from the front and left side. (f)(g) The result before labelling, seen from the front and right. (h)(i) The result using the gray code method, seen from the front and left.

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 major challenge for 3D reconstruction. According to Fig. 5, the detailed structures are reconstructed. The errors are mainly caused by highlights. Because of the highlights, wrong correspondences are introduced, leading the results to be poor. However, the results using our method are still much better than that using the gray code method. Another experiment with a big vase is conducted and the results are shown in the supplemental material. The big vase also has detailed structures on the surface, such as bamboos and leaves. Our results illustrate that the proposed method can reconstruct details on the surface, and the results are better than those using the gray code method.

The plastic cup shown in the supplemental material is quite challenging to be reconstructed because it has two layers. The second layer (the inner one) has strong reflections, which 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. 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, the big holes come from wrong correspondences. However, comparing to the results using the gray code method, our method is much better.

The reconstructed results of a plastic bottle with a green dishwashing liquid inside are also shown in the supplemental material. The dishwashing liquid is transparent and since it has a different refraction index from the plastic bottle, re-

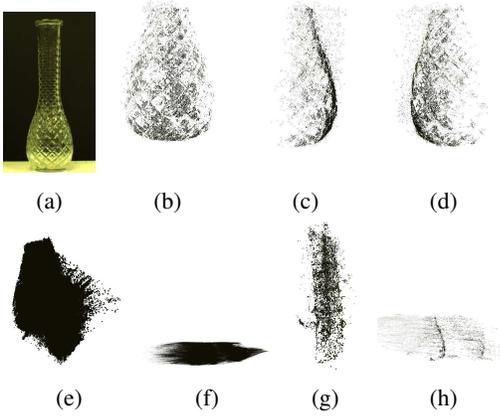


Figure 5: Small vase. (a) The object. (b)(c)(d) The reconstruction result using our method, seen from the front, left side, and right side. (e)(f) The result before labelling, seen from the front and left. (g)(h) The result using the gray code method, seen from the front and left.

fraction and reflection happen at the interface between the bottle and the liquid. For the results using our 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. According to the results shown in the supplemental material, the proposed method has better results than those using the compared method.

Our method can be applied to a large range of objects. Not only for opaque objects and transparent objects, but also for specular objects with anisotropic surfaces, our method can produce quite acceptable results. In the supplemental material, we present the reconstructed results of an anisotropic metal cup. We can easily observe the holes in the middle and on the sides of the results using the gray code method. This is because the surface reflects light anisotropically, and the intensities of these reflections are wrongly interpreted when finding the correspondences. Comparing to the results of the gray code method, our results have much smaller holes and smoother reconstructed surface.

4.2. Quantitative results

Two very challenging objects are used for the quantitative results. The first one is the star trophy shown in Fig. 3. The second is the cone trophy (Fig. 4), with multiple faces.

Eq. 5 defines the root mean square (RMS) error of the correspondences of the results of our method. (x, y) denotes the point in the patterns that has a corresponding pixel in the captured images, and the 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 our 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 defined similarly. According to Eq. 5, 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 ignored in the comparison.

$$C_{F_RMS_error} = \sqrt{\frac{1}{N_F} \sum_{x=1, y=1}^{x=1024, y=768} (C_F(x, y) - C_T(x, y))^2} \quad (5)$$

In addition to the RMS errors, another way to illustrate the quantitative results is shown in Eq. 6, which are used to show the “score” of the frequency-based method and the gray code method. In Eq. 6, 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 larger 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.

$$Score_{F_correspondences} = \frac{\frac{N_F}{N_{all}}}{C_{F_RMS_error}} \quad (6)$$

In addition to comparing correspondences, the distances from the reconstructed points to the camera center are also compared. Eq. 7 defines the root mean square error of the distances for the results of our method. The notations are similar to those of Eq. 5 and Eq. 6. $D_F(x, y)$ denotes the distance from the reconstructed point on the surface to the camera center using our method.

$$D_{F_RMS_error} = \sqrt{\frac{1}{N_F} \sum_{x=1, y=1}^{x=1024, y=768} (D_F(x, y) - D_T(x, y))^2} \quad (7)$$

$$Score_{F_distances} = \frac{\frac{N_F}{N_{all}}}{D_{F_RMS_error}} \quad (8)$$

	Our method	Gray code
Number of reconstructed points	15580	16301
RMS error of the correspondences	4.0570	17.0101
Score based on correspondences	0.1130	0.0282
RMS error of the distances	102.1415	856.9114
Score based on distances	0.0045	5.5991×10^{-4}

Table 1: Comparison between our method and the gray code method for star trophy reconstruction.

Table 1 shows the comparison results of the star trophy reconstruction results using our 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 our method. For our method, the RMS error is not as small as expected. The reason is because around the edges 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 the results in 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 2 shows the quantitative results of the cone trophy reconstruction using our 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 our 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 our method are better than that of the gray code method.

5. Discussions

There are some significant differences between our method and the phase shifting methods. The major difference is that the phase shifting methods can only obtain one correspondence for each image pixel, while our method can

	Our method	Gray code
Number of reconstructed points	26799	21900
RMS error of the correspondences	3.2900	4.0794
Score based on correspondences	0.2683	0.1768
RMS error of the distances	45.9847	43.2997
Score based on distances	0.0192	0.0167

Table 2: Comparison between our method and the gray code method for cone trophy reconstruction.

obtained multiple. The first step of our method applies the frequency-based method to get all possible correspondences for each pixel. The second step described in Sec. 3.3, which is the major contribution of our method, extracts the layer that the user wants. In our paper, we only show the results of extracting the layer that is closest to the camera. However, the energy function can be modified to accommodate for multiple layer extraction. For example, it can be modified to obtain the farthest layer which is the background. Moreover, after we obtained the first layer, this layer can be removed and by running our method again we can obtain the second layer of the objects if there are multiple transparent layers in the objects. To sum up, our method can obtain multiple layers of an object because of the second step which is our major contribution, while phase shifting methods can obtain only one correspondence for each image pixel.

Another advantage of our method is that it works on transparent objects. Using one of the most advanced phase shifting methods [9] as an example, it is noteworthy that none of their experimental objects is completely transparent, to be more specific, only translucent. To reconstruct the 3D shape of completely transparent objects is much more difficult because the light reflected by the object surface is extremely weak. When the reflected light is strongly corrupted by the light reflected from the background, the phase shifting method would definitely fail.

Our method is based on the method presented in [24]. The major different between our method and [24] is the labelling method presented in Sec. 3.3. First of all, the frequency-based method is an essential step for our method because it provides all possible correspondences for each image pixel, where these correspondences can be from the object surface or from the environment. After that, our method extracts the first layer of the object. Furthermore, our method can be modified to extract all the layers of a multiple-layer transparent object.

6. Conclusions and Further Work

In this paper, a new 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 proposed method is based on frequency analysis, which is not effected by noise, the method is 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.

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, our method produces encouraging results. However, for objects with high curvatures or highlighted points, the results of our method are not as good as expected.

Future work can focus on two directions. One is to accelerate the image capturing procedure. For our method, it takes about 33 minutes to capture all the 1350 images. We can use a camcorder to capture the images as in [24]. The other direction is to use a turntable to reconstruct the whole objects from different angles, which may address the issue of highlights.

7. Acknowledgements

The authors would like to thank NSERC, AITF and the University of Alberta for the financial support and the anonymous reviewers for their valuable comments.

References

- [1] B. Atcheson and W. Heidrich. Non-Parametric Acquisition of Near-Dirac Pixel Correspondences. In *VISAPP*, pages 247–254, Rome, Italy, 2012.
- [2] B. Atcheson, I. Ihrke, D. Bradley, W. Heidrich, M. Magnor, and H.-P. Seidel. Imaging and 3d tomographic reconstruction of time-varying inhomogeneous refractive index fields. In *Int. Conf. on Computer Graphics and Interactive Techniques*, 2007.
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*, 23(11):1222–1239, 2001.
- [4] X. Chen, Y. Shen, and Y.-H. Yang. Background estimation using graph cuts and inpainting. In *Graphics Interface*, pages 97–103, 2010.
- [5] Y. Ding, F. Li, Y. Ji, and J. Yu. Dynamic fluid surface acquisition using a camera array. In *ICCV*, pages 2478–2485, 2011.
- [6] G. Eren, O. Aubreton, F. Meriaudeau, L. Secades, D. Fofi, A. T. Naskali, F. Truchetet, and A. Ercil. Scanning from heating: 3d shape estimation of transparent objects from local surface heating. *Optics Express*, 17(14):11457–11468, 2009.
- [7] J. Geng. Structured-light 3d surface imaging: a tutorial. *Adv. Opt. Photon.*, 3(2):128–160, 2011.
- [8] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. Narasimhan. Structured light 3d scanning in the presence of global illumination. In *CVPR*, pages 713–720, 2011.
- [9] M. Gupta and S. Nayar. Micro phase shifting. In *CVPR*, pages 1–8, Jun 2012.
- [10] M. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. Lensch. Fluorescent immersion range scanning. *ACM Trans. Graph.*, 27(3):87:1–87:10, 2008.
- [11] Y. Ji, J. Ye, and J. Yu. Reconstructing gas flows using light-path approximation. In *CVPR*, pages 2507–2514, 2013.
- [12] K. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. *IJCV*, 76(1):13–29, Jan. 2008.
- [13] D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Polarization-based transparent surface modeling from two views. In *ICCV*, pages 1381–1386, 2003.
- [14] N. Morris and K. Kutulakos. Dynamic refraction stereo. In *ICCV*, pages 1573–1580, 2005.
- [15] N. Morris and K. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *ICCV07*, pages 1–8, 2007.
- [16] H. Murase. Surface shape reconstruction of an undulating transparent object. In *ICCV*, pages 313–317, 1990.
- [17] H. Murase. Surface shape reconstruction of a non-rigid transparent object using refraction and motion. *IEEE Trans. on PAMI*, 10(14):1045–1052, 1992.
- [18] D. Nehab, T. Weyrich, and S. Rusinkiewicz. Dense 3d reconstruction from specular consistency. In *CVPR*, pages 1–8, 2008.
- [19] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. on PAMI*, 30(6):1068–1080, 2008.
- [20] B. Trifonov, D. Bradley, and W. Heidrich. Tomographic reconstruction of transparent objects. In *Proc. Eurographics Symposium on Rendering*, pages 51–60, 2006.
- [21] M.-J. Tsai and C.-C. Hung. Development of a high-precision surface metrology system using structured light projection. *Measurement*, 38(3):236 – 247, 2005.
- [22] J. Wang and K. Dana. Relief texture from specularities. *IEEE Trans. on PAMI*, 28(3):446–457, 2006.
- [23] G. Wetzstein, D. Roodnick, W. Heidrich, and R. Raskar. Refractive shape from light field distortion. In *ICCV*, pages 1180–1186, 2011.
- [24] J. Zhu and Y.-H. Yang. Frequency-based environment matting. In *Pacific Conference on Computer Graphics and Applications*, pages 402–410, 2004.