Lighting geometry aware environment matting and 3D reconstruction

Duan, Qi

2012

Duan, Q. (2012). Lighting geometry aware environment matting and 3D reconstruction.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/54835>

<https://doi.org/10.32657/10356/54835>

NANYANG
TECHNOLOGICAL
UNIVERSITY

Lighting Geometry Aware Environment Matting and 3D Reconstruction

A thesis submitted
for a degree of Doctor of Philosophy
by

Duan Qi

School of Computer Engineering
Nanyang Technological University

August 2012

Abstract

Lighting geometry, which refers to the relationship between lighting conditions, object geometry or appearance and various lighting interaction phenomena, is a fundamental issue in many computer graphics and computer vision problems. The challenges of lighting geometry related problems lie in the complex interactions between environment lights and objects, the diversity of lighting sources, the high computational cost and so on. This thesis investigates two specific lighting geometry related topics: environment matting and 3D reconstruction, where the former is with controllable lighting conditions and the latter is with general unknown illumination conditions. Our goal is to design and develop some effective and efficient algorithms which exploit lighting geometry property to improve the performance of the existing algorithms.

First of all, considering that the state-of-the-art real-time environment matting and compositing method is short of flexibility, in the sense that it has to repeat the entire complex matte acquisition process if the distance between the object and the background is different from that in the acquisition stage, and also lacks accuracy, in the sense that it can only remove noises but not errors, we introduce the concept of refractive vector and propose to use a refractive vector field as a new representation for environment matte. Such refractive vector field provides great flexibility for transparent object environment matting and compositing. Particularly, with only one process of the matte acquisition and the refractive vector field extraction, we are able to composite the transparent object into an arbitrary background at any distance. Furthermore, we introduce novel light vector field fitting algorithms to simultaneously remove both noises and errors contained in the extracted matte data. Experimental results show that our method is less sensitive to artifacts and can generate perceptually good composition results for more general scenarios.

Second, considering that the existing high-quality environment matting methods usually require the capturing of a few thousand sample images and spends a few hours in data acquisition, we propose a novel environment matting algorithm to capture and extract the environment matte data effectively and efficiently. Particularly, the recently developed compressive sensing theory is incorporated to reformulate the environment matting problem and simplify the data acquisition process. In addition, taking into account the special properties of light refraction and reflection effects of transparent objects, two advanced priors, group clustering and Gaussian priors, as well as other basic constraints are introduced during the matte data recovery process to combat with the limited image samples, suppress the effects of the measurement noise resulted from data acquisition, and faithfully recover the sparse environment matte data. Compared with most of the existing environment matting methods, our algorithm significantly simplifies and accelerates the environment matting extraction process while still achieving high-accuracy composition results.

Finally, we consider the problem of high-quality 3D reconstruction under unknown illumination using the joint multi-view stereo (MVS) and photometric stereo (PS) technique. We take into account the property of lighting geometry and propose to use total variation term to constrain the light function recovery. Our algorithm can refine the 3D object model and recover the lighting conditions simultaneously. Comparing with many previous methods, our method can provide a significant computational saving and is compatible with traditional MVS methods without the need for extra images.

Acknowledgments

It is a pleasure to thank the many people who made this thesis possible.

First, I would like to express my deep and sincere gratitude to my supervisor Prof. Jianfei Cai for his patience, motivation, immense knowledge and continuous support of my Ph.D study and research. His guidance and encouragement have been of great value for me, which helped me in all the time of my research and writing of this thesis.

Besides, I would like to express my sincere thanks to my co-supervisor Prof. Jianmin Zheng, who is one of the most descent researchers in the related areas. During my PhD career, he inspires me a lot on innovative research ideas with his profound knowledge and amazing wisdom. The discussion with Prof. Zheng would be the most memorable moments during my research career.

Third, I would like to thank my colleagues, especially Dr. Juyong Zhang and Dr. Chunlin Wu, Dr. Shaoting Zhang for your stimulating discussions and valuable advices. You guys shed lights into my research. Meanwhile, many thanks to my relatively old friends, Dr. Lixin Duan, Zhaoyao Zhan, Dr. Wenxian Yang, Haifeng Lu, Tiantian Guo, Ming Yang, Shengbo Yang, Xiao Teng, Zhengxiang Wang, Shenghua Gao, Da Zhang, Xiangang Cheng, Dr.Feng Zhong, Dr. Min Wu, Dr. Kun Zhu for their kind support and assistance. I really enjoy the time spent with every one of you in Singapore and I will always cherish this four years period.

At last, I would like to thank my father, Bairun Duan, for giving birth to me and supporting me spiritually throughout my life. No matter where you are, you are always there with me. Love and miss you forever! Also I would like to thank my mother, Jiansheng Xie and my wife, Dr. Min Tang for your kindly support and love to me. Without you, I could not imagine how can I get through those hard and frustrating days. I pray that you both are healthy and happy forever. I love you.

Contents

Abstract	i
Acknowledgments	iii
List of Figures	vii
List of Tables	xi
1 Introduction	1
1.1 Background	1
1.2 Objective and Research Scope	4
1.3 Summary of Contributions	6
1.4 Thesis Organizations	7
2 Literature Review	8
2.1 Environment Matting and Composition Techniques	8
2.1.1 Image-Based Environment Matting and Compositing	8
2.1.2 Model-based Environment Matting and Compositing	11
2.1.3 Other Related Researches	12
2.2 Compressive Sensing Theory and Applications	13
2.3 Image Based 3D Reconstruction	14
2.3.1 Multi-view stereo reconstruction	14
2.3.2 Photometric stereo reconstruction	15
2.3.3 Combine MVS and PS methods	16
3 Flexible and Accurate Real-Time Environment Matting and Compositing	19
3.1 Motivation and Our Approach	19

3.2	Refractive Vector Field For Environment Matting	21
3.2.1	Refractive Vector Calculation	23
3.2.2	Refractive Vector Field Extraction	25
3.3	Piecewise Refractive Vector Field Fitting	25
3.3.1	Region segmentation	27
3.3.2	B-spline fitting	27
3.3.3	Experiments	30
3.3.4	Limitations	37
4	Compressive High Quality Environment Matting	42
4.1	Motivation and Our Approach	42
4.2	Environment Matting Model and Properties	43
4.2.1	Property of light transport vector	45
4.3	Solving Environment Matting Problem	46
4.3.1	Recovery of camera compensation and light absorption index . . .	46
4.3.2	Light transport vector recovery	48
4.4	Experiments	52
4.4.1	Implementation of light transport vector recovery	52
4.4.2	Visual results of environment composition	53
4.4.3	More comparisons	56
4.4.4	Comparison with compressive image relighting	59
5	Joint Mesh Refinement and Light Function Recovery for High-quality 3D Reconstruction Under Unknown Illumination	61
5.1	Motivation and Our Approach	61
5.2	Lighting geometry property	62
5.2.1	Lambertian lighting model	63
5.2.2	Intensity constraints	63
5.2.3	Piece-wise constant surface light	65
5.3	Optimization Framework for Joint Mesh Refinement and Lighting Recovery	66
5.3.1	Problem Formulation	66
5.3.2	Proposed Optimization Solver	67

5.3.3	Jacobian matrix construction	69
5.4	Experiments	70
5.4.1	Visual results	71
5.4.2	Computational cost	75
5.4.3	Convergence	81
5.4.4	Limitations	82
6	Conclusion and Future Works	83
6.1	Conclusion	83
6.2	Future Directions	85
	References	87
	List of Publications	102

List of Figures

3.1	Illustration of a refractive light route.	23
3.2	Computing a refractive vector shown in a dash line.	24
3.3	Artifacts in environment matting.	26
3.4	Two backgrounds used in the experiments.	31
3.5	Compositing results of a synthetic transparent glass model with two different backgrounds.	33
3.6	Compositing results of a real captured transparent goblet whose extracted matte data contain severe artifacts.	34
3.7	Compositing results of the transparent goblet captured in a different orientation. The extracted environment matte data also contain severe artifacts.	35
3.8	Compositing results of a real captured transparent bowl, for which the extracted environment matte data are quite clean.	36
3.9	Compositing results with a fruit background placed at different distances. In this example, as the background plane moving from near to far, the shape of the refracted grape shown in the center sphere part changes drastically, which are correctly simulated using our proposed algorithm while the previous scaling method fails to recover.	38
3.10	Compositing results with a text background placed at different distances. Similar to Fig. 3.9, as the background plane moves, the refracted text pattern seen through the transparent object changes a lot. Apparently, our proposed algorithm can give more accurate results than the previous method.	39

3.11	An example for which the proposed method does not work well. Here the dragon object contains very much subtle surface geometry that the segmentation method fails to recover, the fitting and composition result is smooth without too much fine refractions comparing to the ground truth.	41
4.1	Simulation of light refraction, reflection and transmission and the corresponding light transport vector. The background resolution is 256×256 and there are three separate parts contributing to certain foreground pixel with oriented Gaussian distribution for the weights of each region. The corresponding light transport vector is sparse with only 2806 non-zero elements out of 65536 elements.	47
4.2	Sample images of a transparent object in front of constant background images with different colors.	48
4.3	Sample images of a transparent object in front of the random background patterns with different resolutions.	49
4.4	Composition results of our proposed methods with and without the Gaussian prior, where the former significantly outperforms the latter in terms of visual quality at the same number of sample images.	55
4.5	Composition results of a 3D glass elephant with 150 sample images. Note that all sample images and ground truth are rendered using 3DS Max.	56
4.6	Our proposed composition result of several complex 3D models. From the results in the first column we can see that due to the complex shape, its refraction and reflection effects are very complicated, but we can still recover such complicated effects efficiently using only 150 sample images and generate very accurate composition results that close to the ground truth.	57
5.1	Illustration of lighting geometry on object surface. No matter what type of environment lights we have, for example, $\{l^k k = 1, \dots, m\}$ from different directions, we can use l_v to represent the combined lighting effects on vertex v . Moreover, we can see that l_v would not change drastically between neighboring vertices, which means it has piece-wise constant property over the surface.	64

5.2	Results of teapot model lit by several directional lights from different directions, from which we can see that its surface illumination condition is piece-wise constant. Here the lighting condition are converted to RGB color for illustration purpose.	66
5.3	Illustration of vertex normal n_a . For a triangular mesh, the normal n_a for vertex v_a is actually calculated according to its one-ring neighboring points.	69
5.4	Examples of real captured sample images of temple and dino objects. . .	72
5.5	Results of the synthesized bunny example. For this bunny example, the input of our framework is a smooth bunny model (shown in the first column) and its rendered images (shown in the second column). Then we can refine the mesh and recover its piece-wise surface lighting function as shown in the third and fifth column, respectively. Compared with the ground truth (column 4 and 6), we can see that our proposed algorithm can generate high-quality mesh with high frequency shape details and recover the lighting function close to the ground truth.	74
5.6	Comparison with Wu's algorithm [126]. For the mesh, our algorithm can refine the initial smooth model and recover mesh surface details similar to [126] but using a more reasonable optimization framework and at a very low computational cost.	75
5.7	Result comparison for the temple model. Here the input smooth model of our proposed algorithm are shown in the first column; the second column shows the vertex intensity of the model which are extracted from real captured images as shown in Fig. 5.4; the reconstructed temple model using conventional MVS algorithm [107] and our proposed algorithm are shown in the last two columns respectively.	76
5.8	Close view of the pillar region of Temple. It can be seen that our proposed algorithm can generate more carved pillar details than [107] as reflected by the photometric cues on the object surface.	77
5.9	Result comparison for the Dino model. The first row shows the reconstructed Dino model using conventional MVS algorithm [107], the second row shows the obtained surface intensity value for each vertex while the last row shows the results of our method.	78

5.10	Comparison of dino head details. Our proposed algorithm can recover the eye and mouth structure for the dino while the conventional MVS algorithm [107] fails.	79
5.11	The calculated surface illumination results for both temple and dino dataset. The surface illumination results are converted into <i>RGB</i> color for illustration purpose. We can see that the illumination results satisfy the piece-wise constant constraint.	80

List of Tables

4.1	Comparisons among different environment matting algorithms with a background image size of $n \times n$ and k denoting the sparsity of the light transport vector.	58
4.2	Average time for the light transport vector recovery using the non-negative dynamic group sparsity recovery algorithm (NN-DGSR) with different settings.	59

Chapter 1

Introduction

1.1 Background

As a very important natural phenomena, lighting always attracts many attentions of experts not only from physics and optics but also from computer graphics and computer vision. In particular, lighting is one of the most fundamental problems in computer graphics and vision research fields, especially playing a very important role in the traditional graphics pipeline¹. On one hand, for the modeling stage of the pipeline or some related computer vision problems, lighting is a lasting and challenging topic and influences a broad range of practical applications, including specular and shadow analysis [106, 85, 92, 113, 93, 34], 3D geometry acquisition [71, 55, 133, 13, 131], object material acquisition [109, 27, 54, 122], object detection and tracking [77, 79, 91, 58, 83, 94], etc. On the other hand, for the rendering part of the pipeline, to model and illustrate various illumination effects accurately, various lighting models are introduced, such as, Lambertian reflectance model [3], Phong shading model [5] and Bidirectional reflectance distribution function (BRDF) [2], etc; meanwhile, to calculate and render such illumination effects realistically, many techniques, for example, ray tracing [8], beam tracing [1], photon mapping [6] and pre-computed radiance transfer (PRT) [7] algorithms, have also

¹Generally speaking, lighting is crucial to the illumination part in graphics pipeline, actually, it is also of great importance in the modeling part since it can help to improve the modeling accuracy [141].

been proposed in recent decades.

In particular, it is obvious that those complicated illumination effects highly depend on the environment lights, object geometry and their interactions. Therefore, we consider to combine these aspects together and propose to investigate the new *lighting geometry* problems which refers to the physical relationship between the lighting conditions, the object geometry or appearance information and the various lighting interaction phenomena that are recorded or reflected in images pixels or real worlds, that is, the underline connections between different stages in general graphics pipeline. Generally speaking, if the object illumination results are already known, any problems that aim to calculate a 3D object model or its surface appearance through lighting conditions, or recover lighting sources information from object shape and appearance knowledge can be classified as the lighting geometry problems.

Such natural illumination effects are very complicated because we need to consider many factors, including light emission, transportation, reflection and refraction and so on. Various lighting geometry topics can be found in computer graphics and vision areas.

- Image-based relighting and rendering (IBRL). Image-based relighting and rendering, which requires to relight and render the real object or scene under arbitrary natural or synthetic illumination conditions without calculating any 3D information, is a typical example of lighting geometry problem and has attracted lots of research interest in computer graphics. In IBRL, complex lighting effects such as subsurface scattering, interreflection, shadowing, mesostructural self-occlusion and other relevant phenomena can be simulated accurately. The main advantage of IBRL is that its computational cost is independent of scene complexity because the rendering process is actually a process of manipulating image pixels, instead of recovering 3D object models and simulating light transport. According to the scene/illumination information acquisition methods, IBRL techniques can be generally classified into

three categories: Reflectance function-based, Basis function-based and Plenoptic function-based algorithm [23].

- Environment matting and compositing. Environment matting and compositing problem is a special case of IBRL which mainly focus on modeling and calculating the complex light refraction and reflection effects of transparent objects. According to the principle of optical path reversibility [18], environment matting problem can be solved by using similar methods of IBRL but under more restrict requirements: the lighting condition is controllable and the transparent object is placed in a dark room to avoid the influence of ambient light and environment light.
- 3D reconstruction. 3D reconstruction is another popular lighting geometry problem, in which the lighting information plays an important role. For one thing, controllable lighting offers great help to recovery the object geometry. As illustrated in photometric stereo (PS) [124, 141], which is a popular 3D reconstruction technique, the object surface normal can be obtained from point shading cues under varying controlled illuminations (e.g. know light directions and intensities) and then the object shape can be recovered by integrating all surface points' normal and position information. Meanwhile, when the lighting condition is uncontrollable, researchers need to focus on how to calculate the general illumination conditions first, and then use such lighting conditions to recover object geometry. Furthermore, lighting can be used to capture the object geometry directly. One example is the structure lighting technique [123], which projects the coded structure lighting patterns onto object surface, records the reflected patterns by one or more cameras and reconstructs the shape of the recorded 3D objects via triangulation of the optical rays corresponding to projector and camera pixels. To recover a high quality 3D model, it requires a large number of sample images and very high computational cost during the model recovery process. Note that the recent developed *Kinect* equipment

can capture the 3D object model in real-time but with much measurement noise and very low resolution.

Therefore, we can see that the lighting geometry problems are of great importance in computer graphics and computer vision. They are worthwhile to be investigated further.

1.2 Objective and Research Scope

Lighting geometry problems have a lot of challenges mainly in the following aspects:

- The interactions between the environment lights and objects are too complex. According to aforementioned analysis of lighting effects, the environment lights can be reflected, refracted and transmitted by the objects while the objects can also emit lights by themselves. In some practical applications, it is too complicated to consider and model all kinds of lighting effects together for the target lighting geometry problem.
- The environment light sources are diverse. In real worlds, there are a wide variety of light sources, such as, sun light, fluorescent lamp, incandescent light, LED light, etc, and different types of light shapes including point light, bar light, ring light, plane light and so on. It is hard to model the illumination conditions of arbitrary unknown environments accurately.
- The computational cost is very high. Since the light interactions and lighting conditions are very complicated and the object geometry is arbitrary, such lighting geometry problems are often of very high dimension and require large computational cost. For example, when the environment lighting conditions are complex, the IBRL technique requires to capture a large number of sample images and build a complicated environment lighting model to approximate the general real lighting condition; to obtain an accurate 3D object model with high-frequency surface

details, a large number of sample images and high computational cost are also unavoidable.

In this thesis, we mainly focus on two specific lighting geometry related topics: environment matting and 3D reconstruction, which are just two typical problems corresponding to different parts of the graphics pipeline. As mentioned before, the former topic belongs to the render stage which not only extract the environment light information but also render the light transmission, reflection and refraction effects of the transparent object while the latter one aims to recover an accurate object 3D model using lighting priors in the modeling part. Our goal is to design and develop some *effective* and *efficient* algorithms which exploit such *lighting geometry* property to improve the performance of current environment matting and compositing techniques and 3D reconstruction algorithms.

Specifically, this thesis aims to tackle the above challenges in the following aspects:

- Improve the flexibility of conventional real-time environment matting and compositing techniques by introducing the new concept of refractive light vector field as the new representation for environment matte, which can better illustrate the essence of light refraction phenomenon; moreover, to improve the accuracy of extracted matte data we propose a novel piecewise vector field fitting algorithm to simultaneously remove both noises and errors contained in the captured matte data;
- Simplify the sample image acquisition process and accelerate the matte recovery process by incorporating the compressive sensing theory into high-quality environment matting extraction problem. In order to suppress the measurement noise and faithfully recover the sparse environment matte data with limited sample images, we propose to use the group clustering and Gaussian priors during the matte extraction process;

- Reduce the computational cost of 3D reconstruction problem under general unknown illumination conditions by investigating the property of the lighting conditions on object surface. According to surface shading cues, we propose a novel optimization framework that can refine the object models and calculate the piecewise constant illumination conditions simultaneously.

1.3 Summary of Contributions

The major contributions of this thesis include:

- We introduce the new concept of refractive light vector and propose to use a refractive light vector field as the new representation for environment matte, which provides great flexibility for environment matting and compositing of transparent objects. Furthermore, we propose a piecewise vector field fitting algorithm to simultaneously remove both noises and errors contained in the extracted matte data, which makes our proposed environment matting method less sensitive to artifacts and suitable for more general sceneries;
- We propose to incorporate the recently developed compressive sensing theory into high-quality environment matting problem to simplify the data acquisition process, reduce the sample image number and the computational cost for environment matte recovery. Taking into account the special properties of light refraction and reflection effects of transparent objects, the group clustering and Gaussian priors are introduced during the matte data recovery process. These two constraints can efficiently combat with the limited sample images, suppress the measurement noise and faithfully recover the sparse environment matte data. Compared with most of the existing methods, our algorithm significantly simplifies and accelerates the environment matting extraction process while still achieving high-accuracy composition results;

- We analyze the property of the surface illumination conditions of Lambertian objects under general environments and propose a piece-wise constant constraint on object surface lighting conditions. Then we introduce a novel optimization framework for 3D model refinement by utilizing such light constraint and photometric cues. Based on piece-wise constant light priors, our optimization framework can refine the object model and recover its light condition simultaneously. Comparing with many previous methods, our method is effective, efficient and compatible with traditional MVS methods without capturing any extra images.

In next section, we introduce the organizations of the thesis. Readers can jump to the part of their particular interest.

1.4 Thesis Organizations

The rest of this thesis is organized as follows. In chapter 2 we give a rough literature review, about the existing environment matting and compositing approaches, the recently developed compressive sensing theory and the current existing 3D object reconstruction algorithms. In chapter 3 we will present the new concept of refractive light vector field and introduce the vector field fitting algorithms to remove the artifacts in real-time environment matte extraction method. In chapter 4 we will introduce our novel compressive environment matting framework and some important priors that can significantly simplify and accelerate the high-quality environment matting extraction process, reduce the computational cost while achieving high-quality compositing results with the limited sample images. In chapter 5 we will present our novel joint mesh refinement and light function recovery framework that can simultaneously refine the object model and recover its surface illumination results. Finally we will summarize this thesis and discuss the direction of our future work.

Chapter 2

Literature Review

In this chapter we give a review about the related works. To begin with, we introduce the environment matting and compositing approaches which consider the light refraction and reflection effects of transparent objects. Then we review the recently developed compressive sensing theory that is to be used in our research, as well as its applications. Finally, we summarize the existing 3D object reconstruction algorithms.

2.1 Environment Matting and Composition Techniques

2.1.1 Image-Based Environment Matting and Compositing

In digital image editing, matting and compositing are two fundamental techniques [105]. Conventional image matting and composition techniques handle opaque objects, which extracts a foreground object of arbitrary shape and its related information from a source image, and the process of compositing places the foreground object over an arbitrary background using the matte to control the contribution. Usually such image matting and compositing techniques can generate fantastic visual effects, but they have difficulty in handling transparent objects which often exhibit refraction and reflection phenomena. To directly fetch the light refraction and reflection information from uncalibrated source images is almost impossible. Environment matting and compositing algorithms are therefore developed, which generalize the conventional matting and compositing by incorpo-

rating the information of how a foreground object refracts and reflects light from the environment.

On the other hand, according to the principle of optical path reversibility [18], environment matting techniques can also be interpreted as its inverse problem—image based relighting which is based on directly measuring surface. Similar to image-based relighting algorithms, to fully recover the refraction and reflection information of a transparent object, the straightforward “brute force” method is to illuminate the object by only one point light at a time and sweep over the whole background image as the raster scanning method. The complexity of such method is $O(n^2)$ where n^2 is the area of the environment around the target object. Such time complexity can be reduced to $O(\log n)$ by hierarchy of progressively finer patterns [57, 52, 97]. Enlighting by these algorithms, Zongker *et al.* [144] first proposed the formula of environment matting as:

$$\begin{aligned} C &= F + (1 - \alpha)B + \Phi \\ &= F + (1 - \alpha)B + \sum_{i=1}^m R_i \mathcal{M}(T_i, A_i) \end{aligned} \quad (\text{Eq. 2.1})$$

where R_i is new reflectance coefficient for certain region A_i located at background texture T_i , and $\mathcal{M}$ is the “texture mapping operator” [144]. Compared to traditional image matting approaches, the additional special refraction and reflection properties of the transparent object are represented by Φ in this model. To solve this environment matting problem and recover Φ value, horizontal and vertical strip patterns are typically used and emitted onto a scene and then images of the transparent object under such patterns are captured, where the structured backgrounds consist of a hierarchy of finer and finer horizontal and vertical square-wave stripes. Finally the environment matte of each foreground point (or pixel) is calculated and represented by a reflection coefficient and a corresponding rectangular area on background.

There are still two shortcomings in this environment matting algorithm. First, if the background is a $k \times k$ pixel grid, $O(\log k)$ images are needed, which makes the capturing

process complicated and time-consuming. Second, the background patterns used in [144] are axis-aligned rectangular patterns, which results in a lot of excessive blurring and aliasing in the compositing results. Later on, Chuang *et al.* [24] proposed two extensions to enhance the usability of the environment matting algorithm of [144]. The first extension aimed to improve the accuracy of the matting results by using 2D oriented Gaussian pattern instead of axis-aligned rectangle to recover light spatial variation and dispersion effects as well as multiple mappings of texture, which can better approximate the bidirectional reflectance distribution function (BRDF) model. The second extension was targeted to achieve fast/real-time environment matting by making certain assumptions and simplifying the matting process to the one with only one picture captured against a special backdrop. The real-time environment matting is more practical and has been widely used in many applications. However, this real-time method is sensitive to artifacts, for which the anisotropic filter [120] was used in [24] to eliminate the artifacts caused by noises.

There also exist some other solutions for environment matting and compositing problems. For example, Wexler *et al.* [121] captured a series of images of the foreground transparent object with a moving background and recovered a clean background image without the foreground object by using some computer vision techniques [61, 110]. After that, a probabilistic model is used to extract the matte by assuming that each background pixel has a certain probability to make contribution to the final color of certain foreground pixel. With some priors the environment matte of the transparent object can be recovered. This algorithm is able to work under uncontrolled conditions without any knowledge of the background patterns, but the final matte result heavily depends on the accuracy of the computed clean background, which requires large number of image samples. To reduce the number of image samples, Peers and Dutré [86] applied wavelet transform used in image compression [11, 31] to decompose an image into a combination

of sets of basis vectors. If proper basis vector sets to represent the background image can be estimated, the number of image samples can be reduced significantly. Similarly, Zhu and Yang [143] introduced a frequency-based method to solve environment matting problem. The background patterns they used consist of signals of different frequencies at different regions. The environment matte data is calculated by finding a mapping between a foreground point and background regions that have the same frequency response after Fourier transform.

In summary, most of the existing image based environment matting algorithms, especially the wavelet method [86] and the frequency-based method [143], require a large number of image samples, which makes the capture process very complex and time-consuming.

2.1.2 Model-based Environment Matting and Compositing

Compared with image-based methods, the model-based environment matting is based on some known object models and can also synthesize the results of transparent objects under arbitrary environment conditions. As we know, there is a similar concept in computer graphics named “environment map” [46, 49], which is usually used to reduce the complexity and time costs when rendering diffusive and specular materials. Some algorithms have been developed to render a target object with refraction and reflection property into new environment. For example, Blinn *et al.* [15] used the environment mapping technique that calculates the environment reflection result. Greene [46] added some shading effects into Blinn’s method. Voorhies *et al.* [116] used texture mapping hardware to accelerate the rendering process of such environment mapping method and achieve an interactive rate. Kay *et al.* [67] also developed an approximated method that can reduce the time complexity to render a refractive object.

In another aspect, the light transportation procedure can ideally be recovered if an accurate geometric model of the object and its refractive indices are available. Then the

result of the transparent object under new background can be rendered by modern ray tracing techniques [41, 103] and can sometimes be achieved at an interactive speed [118]. However, the final rendering result are often unsatisfactory due to the following limitations. First, it is difficult to obtain an accurate geometrical shape of the transparent object using traditional 3D laser scanning equipments. Second, the material of the object may not be homogeneous everywhere and thus the transportation route changed inside the transparent object is complex which can not be accurately simulated using the state-of-the-art algorithms.

2.1.3 Other Related Researches

There are also some interesting researches on transparent objects which combine environment matting with other techniques. For example, Matusik *et al.* [75] built a complex light stage system, which consists of a turntable, a set of cameras and lights, and monitors. The system can capture environment matting under multiple viewpoints and reconstruct 3D visual hull of transparent objects and their surface reflectance field. In this way, a transparent object can be synthesized under new viewpoints, background and illuminations. Moshe *et al.* [14] used a moving camera to capture a transparent object and tracked some feature points inside the object. Then, based on the pre-calculated camera motion and the refracted light direction, an approximate shape of the object can be estimated. Agarwal *et al.* [9] used the optical flow method to track and simulate the refractive effect of a transparent object from a video sequence of the object in front of a moving background. All these methods involve 3D reconstruction of a transparent object and their computational complexity is very high. Other studies on the 3D reconstruction of transparent (or specular) objects include [78, 134, 80, 82, 81, 69, 70, 135].

2.2 Compressive Sensing Theory and Applications

The recently developed compressive sensing theory (CS) demonstrates an efficient way to reconstruct sparse signal with a small number of measurements. Suppose there is an unknown vector $x \in \mathbb{R}^n$ and we have its measurement result $y = \Phi x$ where Φ is a $m \times n$ measurement matrix. If $m < n$, it is an ill conditioned problem to recover a general x from y . But if x is a k -sparse signal with at most k ($k \ll n$) non-zero elements, we can recover the signal x by solving the following optimization problem:

$$\operatorname{argmin} \|x\|_0, \quad \text{subject to } y = \Phi x \quad (\text{Eq. 2.2})$$

which is an NP-hard problem and can only be solved by the greedy algorithm. The main drawbacks of the greedy methods are that the reconstructed result is often suboptimal and it may be trapped into local optimal if the initial guess is not good. Furthermore, obviously it is very sensitive to measurement noise.

According to the compressive sensing theory [33, 21, 22], if the measurement matrix Φ satisfies the restricted isometry property (RIP) [22], the above NP-hard problem is equivalent to its convex relaxation

$$\operatorname{argmin} \|x\|_1 \quad \text{subject to } y = \Phi x \quad (\text{Eq. 2.3})$$

which can guarantee a global optimal solution and can be reliably solved in polynomial time with only $\mathcal{O}(k \log k)$ non-adaptive measurements.

To be noticed that measurement matrix Φ with RIP means that for any k -sparse vector x and ϵ , the following inequality holds:

$$(1 - \epsilon)\|x\|_2 \leq \|\Phi x\|_2 \leq (1 + \epsilon)\|x\|_2. \quad (\text{Eq. 2.4})$$

This inequality indicates that the eigenvalues of any submatrix of Φ are very close to 1. No matter where the non-zero elements are in the sparse vector, they will be equally

sampled, which guarantees the recovery of the sparse signal with insufficient number of measurements. The popular matrices satisfying the RIP include Gaussian matrices, Bernoulli matrices and partial Fourier matrices [12].

Due to the excellent performance of compressive sensing in recovering sparse signals, it has been widely used in many applications, such as signal processing [45, 90], image reconstruction [117, 36], medical image analysis [73, 50], shape prior modeling [142], and so on. Also in many computer graphics research fields, compressive sensing helps to reduce the time complexity and accelerate the existing algorithms, for example, simulate light transport in volumetric media [47], accelerate existing ray tracing and rendering methods [101, 102] and reduce the complexity for image based relighting [88] and dual photography [100].

2.3 Image Based 3D Reconstruction

To reconstruct the 3D model of an object, multi-view stereo (MVS) and photometric stereo (PS) are two fundamental and widely used techniques in computer vision. Both of them aim to recover the object model from images but in different ways and have different features accordingly.

2.3.1 Multi-view stereo reconstruction

Multi-view stereo (MVS) reconstruction is an important problem in computer vision. Its goal is “*given a set of input photographs, how to estimate a 3D shape that would generate the same photographs, assuming constant material, viewpoints and lighting conditions*”. In the last decade, lots of attentions have been attracted in this problem and various techniques have been developed that can achieve a high degree of reconstruction accuracy. According to [98], the MVS methods can be classified into four categories: volumetric methods [42, 111, 115, 104, 107], surface evolution methods [140, 40, 30, 68, 56, 112, 137],

depth map fusion methods [42, 76, 20, 19, 72] and surface region growing methods [48, 43, 39]. Generally speaking, the volumetric methods first give initial volume of the whole scene, split scene volume into several voxels and then determine the occupancy of each voxel to generate the final object model; the surface evolution methods first give a rough initial shape (e.g. sphere) for the object, then evolve the mesh from rough to fine according to shape consistency or other related constraint between multi-view images; the depth map fusion methods would calculate the depth map result under each viewpoint and then merge those depth map together to generate the 3D object model; the surface growing methods first find several seed surfaces with the matching 3D surface points through triangulation method, expand those seed surfaces gradually and then calculate the 3D object shape from all result seed surfaces.

For MVS methods, its essential part is to find the point matches from multi-view images and calculate its 3D coordinates through triangulation method, which means the point matching between images is very crucial to the reconstruction model. So far, it is still difficult to establish such accurate pixel correspondence for all surface points between multi-view images, especially for object with less texture or smoothly varying surface, where the MVS methods have to adopt interpolation or approximation techniques in uncertain surface regions. From the MVS evaluation benchmark [98] we can see that even the-state-of-the-art MVS algorithms can only recover rough object models with missing high-frequency shape details.

2.3.2 Photometric stereo reconstruction

Photometric Stereo (PS) is also a popular 3D reconstruction technique which can generate very high quality surface models. The basic idea of PS is that the object shape model can be obtained by integrating all surface points' normal information with position information where the normal information can be recovered from point shading cues under

varying controlled illuminations. Many PS algorithms have been proposed to recover the surface normal and model. For example, PS under complex illumination situation was demonstrated in [13]; Hertzmann *et al.* [55] introduced the example based idea into PS, which calculated the geometry of object as well as its reflectance property according to the illuminated results of some example objects (with similar material and known geometry) placed together in the same scene; Tai-Pang *et al.* modeled all surface point as a Markov Random Field, and reformulated and solved the PS problem using some popular optimization algorithms [133, 132]; Hernández *et al.* [53] used colored lights to measure surface normals on deforming objects; Sun *et al.* [108] used multi-light photometric stereo technique to recover non-Lambertian object surfaces with shadow and specular regions; Lun *et al.* [131] can recover both surface normal and shadow and specular information by casting the surface normal recovery problem as a low rank matrix completion problem. While most of the above algorithms only work for object with simple reflectance property (Lambertian reflectance model), Alldrin *et al.* [10] and Goldman *et al.* [44] proposed to recover the shape, surface normal and even spatially varying BRDF information of object with very complicated reflectance properties.

It is obvious that PS algorithms have at least two drawbacks: 1) it is difficult to handle moving or deforming object because the scene should keep still when the illumination conditions change; 2) compared with MVS, it is actually a 2.5D reconstruction algorithm because all the input images are from a fixed viewpoint.

2.3.3 Combine MVS and PS methods

From the above review we can see that MVS algorithm can generate a rough 3D object model but often lose high-frequency shape features, while PS is able to recover accurate object surface details but only under a fixed viewpoint. Therefore, some interesting work has been done trying to combine these two techniques together for object reconstruction

purpose, like [63, 64, 65, 66, 62, 37, 125, 126, 137]. In the early stage, Fua *et al.* [38] minimized the error function of shading, surface smoothness as well as stereo constraint together to refine a coarse model of multi-view stereo method, which can reconstruct object with smooth varying albedo properties. Samaras *et al.* [96] calculated both the object shape and piece-wise constant reflectance albedo using multi-view images in fixed illumination. Jin *et al.* [63, 64, 65, 62] introduced variational frameworks into this area that combine various MVS or PS constraints together, and they can estimate the shape of Lambertian objects, surface albedo and the lighting condition through surface evolution and variational minimization methods. Also, Joshi *et al.* [66] proposed to merge the depth map and surface normal field together to reconstruct object model while Hernández *et al.* [37] chose a two-step scheme similar to [84] that first recovered the surface normal direction and then refined the initial MVS model accordingly. Although the above two methods produce good 3D models, they both require a very large number of sample images in different viewpoints and varying lighting conditions in each view angle. Moreover, they both assume that a single point light source is used in a dark room, and cannot handle general situations. Yoshiyasu *et al.* [137] introduced a new topology adaptive mesh evolution method to recover object models by evolving the initial mesh so as to satisfy MVS boundary and PS normal constraints. Compared with [66, 37], the method in [137] reduces the number of required sample images but still assumes a single light source with known position and direction in a dark environment. Meanwhile, Wu *et al.* [126] used sphere harmonic functions to model general illumination conditions and refined the object model through shading cues accordingly, where their output model contains much surface details. However, there are still some drawbacks in [126]. On one hand, their approach is of very high computational cost because the sphere harmonic light function recovery process is very complicated and the final mesh model is refined vertex by vertex due to the nonlinearity of the energy function. On the other hand, their sphere harmonic

illumination results, which are optimized using the initial MVS mesh model, are not very accurate enough actually. In another aspect, for reconstructing dynamic objects, Vlasic *et al.* [114] and Wu *et al.* [125] both built very complicated equipments with multiple lights and cameras, which can be used to compute accurate surface normal field under each camera view and merge these data together using MVS techniques. The difference between them is that the former only uses visual hull constraints for a real-time merging and reconstruction while the latter uses sphere harmonic function to represent the uncalibrated lighting conditions and calculates full points correspondences for accurate 3D reconstruction purpose. Note that most of the above methods only consider Lambertian objects. There are also some algorithms proposed to handle non-Lambertian objects. For example, Yu *et al.* [138, 139] proposed to reconstruct non-Lambertian object models in multi-view using distant light sources. Variational framework was also adopted in [136] that combines both MVS and PS cues to generate 3D models with dichromatic surfaces and Phong shading model, while its drawbacks lie in the requirement of known lighting directions and the smooth reconstruction results because the variational model is likely to be trapped in local minima.

In summary, the PS methods can improve the accuracy of MVS reconstruction techniques but still have some drawbacks, such as, requiring a large number of sample images, high computational cost, known illumination conditions and so on.

Chapter 3

Flexible and Accurate Real-Time Environment Matting and Compositing

In this chapter, we propose our novel algorithms to increase the flexibility and accuracy of existing real-time environment matting and compositing methods, which makes our proposed real-time environment matting method less sensitive to artifacts and suitable for more general sceneries.

3.1 Motivation and Our Approach

The concept of environment matting and compositing was first proposed by Zongker *et al.* [144]. They demonstrated the ability of environment matting and compositing in capturing and rendering the effects of reflection, refraction, and scatter of light from environment, which greatly improves the visual reality of resulting images. Since then, some environment matting and compositing techniques have been proposed [24, 75, 86, 121, 143]. Basically, these researches were conducted towards two goals: higher accuracy and real-time processing. Surprisingly, none of them has taken the distance between the foreground object and the background into consideration during the matting and compositing processes. Specifically, the existing environment matting and compositing techniques require the distance between the object and the background to remain the same during the matte extraction and the composition processes. If the distance is changed, the

matte has to be re-extracted. This is very inconvenient since in practice, when a matte is extracted from one environment and used for composition in a new environment, the distance between the object and the new background could be different from that in the first environment. Perhaps one could scale the background to an appropriate size and then use the scaled background as the input for [24] to handle the distance issue. Unfortunately, this simple extension does not work very well, as shown in Fig. 3.9 and 3.10. Moreover, the extracted environment matte data often contains noises that are some small unwanted perturbations and errors that are salient outliers in a certain local region. Although filtering approaches have been employed in the existing real-time environment matting and compositing methods [24] to remove noises and have yielded acceptable composition results, they are incapable of removing errors.

To overcome these limitations of the existing real-time environment matting and compositing methods, in this chapter, we first introduce the concept of refractive vector and propose to use a refractive vector field as the new representation of environment matte. Similar to the work in [24], to extract the initial environment matte data, here we assume that the transparent object is colorless and specularly refractive. This will simplify our discussion in elaborating our new idea and enable us to focus on simulating refraction property, which is the most typical feature of transparent objects. Note that under such assumption about the transparent object [24], for each point (or pixel) on the foreground transparent object, there is a ray originating from the background and hitting on the transparent object. It then enters the object and eventually exits the object at the point towards the camera. We reverse all these rays to form a refractive vector field. Such a refractive vector field can be regarded as an attribute of the transparent object and is independent of the background. Once the refractive vector field is obtained, it is possible to place the object over a new background with a changing distance to the object (see Fig. 3.9 and 3.10). Second, we propose a piecewise vector field fitting algorithm to

refine the refractive vector field, which can simultaneously eliminate both noises and errors in the extracted matte data. Experimental results show that, compared with the existing real-time environment matting and compositing methods [24], our algorithm is less sensitive to artifacts and can generate perceptually good composition results in more general scenarios.

The contributions of this work are twofold. First, although it is not new to study the light transportation route passing through a transparent medium, to the best of our knowledge, this is the first work to explicitly use the last transportation route as the matte data for environment matting. Secondly, our contribution lies in the proposed two refractive vector field fitting algorithms, which takes the characteristics of the refractive vector field into consideration and eliminates both noises and errors effectively.

3.2 Refractive Vector Field For Environment Matting

The environment matting equation introduced in [144] is

$$C = F + (1 - \alpha)B + \sum_{i=1}^m \rho_i \mathcal{M}(T_i, A_i) \quad (\text{Eq. 3.1})$$

where C , F and B are the colors of image, foreground, background respectively, α is the opacity value of the foreground object, ρ_i is the reflectance that describes the contribution of light emanating from the environment to the object, $\mathcal{M}$ is the “texture-mapping operator” calculating the average color of region A_i in background texture T_i , and m is the number of the structured background textures. The region A_i is specified by the center $c = (c_x, c_y)$ and the size $w = (w_x, w_y)$ where w_x and w_y are the dimensions in x - and y -directions.

The approach in [24] simplifies the environment matting equation by making some assumptions with a focus on the refraction effects of the transparent object. With the assumption that the object is colorless and purely specularly refractive, the blending

effects between the foreground object color F and the background plane B (that is the $F + (1 - \alpha)B$ term in (Eq. 3.1)) can be ignored. While only one specially designed background is used in the matte acquisition process, the summation and the script i can be dropped and $\mathcal{M}(T, A) \approx T(c)$. Thus the environment matting equation is simplified as

$$C = \rho \cdot T(c) \quad (\text{Eq. 3.2})$$

which implies that the final captured color seen from one point of the transparent object is totally depend on only one point (or region) of the background, as illustrated in Fig. 3.1. Based on (Eq. 3.2), only c and ρ are extracted, which are treated as the matte data [24]. This approach is simple and it can capture environment matting data in real time with only one-shot picture of the object against a specially designed backdrop. Essentially, it establishes a point-to-point (or region) mapping between the object and the background plane. However, when the distance between the object and the background plane is changed, such mapping has to be rebuilt, which means that we have to repeat the complicated and time-consuming data acquisition process.

In this research, we propose a new idea to solve the above problem, which is to calculate a new geometry element (rather than c) as the target matte. Let us consider the ray originating from the camera and casting to a point on a transparent object (see Fig. 3.1), which enters the object and eventually exits the object towards the background. This ray route is the reverse of an actual light originating from the background and finally entering into the camera after refraction. Here we are interested in the last transportation route (from the object to the background) of the ray, which we call a *refractive vector*. Note that by this definition the refractive vector can be specified by a point and a direction. It is not simply a free vector. For each point (or pixel) on the foreground transparent object, there exists a corresponding refractive vector. All these refractive vectors form a refractive vector field, which we choose as our new representation of

Figure 3.1: Illustration of a refractive light route.

environment matte. The refractive vector field can be regarded as an attribute of the foreground transparent object since it depends on the object and is independent of the background. In the following we describe how to represent, compute and extract the refractive vector field.

3.2.1 Refractive Vector Calculation

Suppose we have two samples which are the shots of a transparent object against the background at two different distances. Denote the distances between the object and the two backgrounds by d_1 and d_2 , respectively. Considering a foreground pixel (x, y) in the two pictures, if the corresponding regions in the first and second backgrounds that contribute to the color of this pixel are centered at coordinates (c_{1x}, c_{1y}) and (c_{2x}, c_{2y}) , then $(c_{2x} - c_{1x}, c_{2y} - c_{1y}, d_2 - d_1)$ is the direction of the refractive vector (see Fig. 3.2).

Figure 3.2: Computing a refractive vector shown in a dash line.

Next we need to choose a point on the refractive vector. For simplicity, we choose the point at which the refractive vector hits on the foreground plane and call it a *starting point*. If we let the foreground plane have a z -coordinate of 0, then by some calculations we obtain the starting point $(x_0, y_0, 0) = (\frac{d_2 c_{1x} - d_1 c_{2x}}{d_2 - d_1}, \frac{d_2 c_{1y} - d_1 c_{2y}}{d_2 - d_1}, 0)$. The refractive vector is thus described by parametric equation as

$$P(t) = (x_0, y_0, 0) + (c_{2x} - c_{1x}, c_{2y} - c_{1y}, d_2 - d_1)t$$

where t is a distance parameter.

In practice, we can let the difference between d_1 and d_2 be fixed (say $d = d_2 - d_1$). Then we only need to store the first two components of the starting point and the direction of the refractive vector, which form a 4D vector $(x_0, y_0, c_{2x} - c_{1x}, c_{2y} - c_{1y})$. We denote this 4D vector by $\mathcal{R}(x, y)$, which is the representation of our proposed refractive vector.

Once the refractive vector is extracted, for a new background with a distance of d_3 to the object, we can easily find the corresponding mapping center (c_{3x}, c_{3y}) in the new background for foreground point (x, y) :

$$(c_{3x}, c_{3y}) = (x_0, y_0) + \frac{d_3}{d}(c_{2x} - c_{1x}, c_{2y} - c_{1y}) \quad (\text{Eq. 3.3})$$

3.2.2 Refractive Vector Field Extraction

Now let us look at how we can extract the refractive vector field of a transparent object. During the data acquisition process, we fix the distance between the camera and the object while placing the background (a monitor) at two different distances. For each distance, we adopt the simplified environment matting model (Eq. 3.2) proposed in [24] to calculate the corresponding mapping center c in the background and ρ for each pixel of the object, which requires only two images, with and without the transparent object in front of the specially designed background (a planar slice through the RGB cube). Once the two mapping centers are found, we can compute the refractive vector using the proposed method introduced in Section 3.2.1. In this way, we extract the entire refractive vector field for the foreground object, which requires only two pairs of images in total.

3.3 Piecewise Refractive Vector Field Fitting

Ideally, with the established refractive vector field, we should be able to composite the transparent object into any new background image. However, the pre-generated vector field usually contains a quite significant number of artifacts, which are brought in from the data acquisition and matte extraction process. Fig. 3.3 shows examples of directly re-using the pre-generated refractive vector field for composition. In the figure, some marks are added to show the direct vectors from the red points on the transparent object to the corresponding mapping points in blue on the background.

3.3.a: Noise effects

3.3.b: Error effects (Marked)

Figure 3.3: Artifacts in environment matting.

Artifacts in the extracted refractive vector field can roughly be classified into noises that are small perturbation existing randomly in the whole refractive vector field and errors that are salient outliers in some local regions and are generated because of some assumptions of the approximate model or processing. Some approaches have been developed to remove such artifacts. For example, Chuang *et al.* [24] used the anisotropic filtering method [89], which can remove noises after several iterations but is not suitable for removing errors. While most environmental matting methods handle those artifacts equally, Duan *et al.* [35] proposed a method to treat them differently. The method identifies errors as the outliers in each horizontal line. A weighed curve fitting method is used to fit the direct vector field in each horizontal line with lower weight assigned to error vectors. The method can obtain a visually satisfactory result but it is limited to some symmetric situations.

In this chapter, we present a piecewise fitting method to remove both noises and

errors of the refractive vector field. The method consists of region segmentation which decomposes the whole foreground area into some regions that have relatively consistent refractive vectors, and B-spline fitting which fits a 4D B-spline surface to the refractive vector field for each region. It is worth pointing out that in this method each region is independently fitted by a B-spline surface and there could be discontinuities around the region boundaries. Some feature-preserving boundary smoothing methods may be used for blending region boundaries. However, since the regions are obtained by the region segmentation process based on the refractive vector variance and the refractive vector field may not be continuous from one region to another, it is not very necessary to make a transition between regions to make the field continuous around the region boundaries. Our experiments also show that adding boundary smoothing does not render apparent visual improvement.

3.3.1 Region segmentation

The purpose of region decomposition is to segment the whole foreground object area into some regions such that each of the regions has consistent refractive vectors. The process of region segmentation is imperative because it affects the effectiveness of the subsequent B-spline surface fitting. In this research, we adopt MeanShift algorithm [25] to classify all refractive vectors and segment the whole refractive vector field into sub-regions based on a metric that considers both the distance of two pixels and the difference of their refractive vectors. To avoid over-segmentation, we also merge some small regions after the MeanShift segmentation by comparing the mean value and variance of the refractive vectors of small regions.

3.3.2 B-spline fitting

Consider a region $\Omega \in \mathbb{R}^2$ which is a part of the foreground area. The extracted refractive vector field over Ω is denoted by $\mathcal{R}^*$, which may contain noises and errors. Thus $\mathcal{R}^*$ can be

written as $\mathcal{R}^* = \mathcal{R} + \mathcal{E} + \mathcal{N}$, where $\mathcal{R}$, $\mathcal{E}$ and $\mathcal{N}$ stand for the ideal refractive vector field, errors and random noises, respectively. Our goal is to recover $\mathcal{R}$, the target refractive vector field that does not have artifacts. This is very difficult or even impossible if an analytical solution is sought. Therefore we consider this artifact removal task as an energy minimization problem, that is to minimize:

$$\sum_{(x,y) \in \Omega} (\|\Delta \mathcal{R}(x,y)\|^2 + \lambda \omega(x,y) \|\mathcal{R}^*(x,y) - \mathcal{R}(x,y)\|^2) \quad (\text{Eq. 3.4})$$

where Δ is the Laplacian operator ($\Delta = \frac{\partial^2}{\partial x^2} + 2\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2}{\partial y^2}$) over the refractive vector field, the first item is for the smoothness of the refractive vector field, the second item is to maintain data fidelity during the minimization process, λ is a trade-off factor (in this work, we set $\lambda = 10$), and ω is a weight for each refractive vector. This is a typical weighted least square fitting [32, 51]. If the weights are appropriately chosen, the solution of (Eq. 3.4) could be a good approximation to the ideal $\mathcal{R}$.

The choice of weight $\omega(x,y)$ is crucial to the final solution. This is because the weight reflects the importance of a refractive vector. As verified in [35], it is better to give the highest weight, a lower weight, and the lowest weight to regular refractive vectors, noise vectors, and error vectors than to treat them equally in [24]. In this way, the error refractive vectors will be ignored in the fitting procedure, and through the energy minimization, the corrupted error refractive vectors can be pulled back towards the regular ones.

In particular, since refractive vectors come from the refraction phenomena, the change of the refractive vectors in a smooth region should be consistent, which is mentioned as the criterion of “smooth change” in [28]. For point (i,j) in a certain smooth region, if the average of the refractive vectors in that region is denoted by $\bar{\mathcal{R}}(i,j)$, we measure the local consistency at (i,j) as

$$L(i,j) = \|\mathcal{R}(i,j) - \bar{\mathcal{R}}(i,j)\|. \quad (\text{Eq. 3.5})$$

The value of $L(i, j)$ can be used to detect whether there exist severe errors in the extracted refractive vector field. If $L(i, j)$ is above a certain threshold (say, $2 \times \|\bar{\mathcal{R}}(i, j)\|$), such a refractive vector is considered to be an error vector to which zero weights should be given. Moreover, considering that transparent objects such as glass are typically of piecewise smooth shape, it is reasonable to assume that the refractive vector field is locally smooth. We measure the smoothness of a refractive vector $\mathcal{R}(x, y)$ at coordinates (i, j) through calculating the difference of the 4D vector $\mathcal{R}(x, y)$ and its neighbors:

$$S(i, j) = \|\mathcal{R}(i, j) - \frac{1}{4} \sum_{(i^*, j^*) \in N_1(i, j)} \mathcal{R}(i^*, j^*)\| \quad (\text{Eq. 3.6})$$

where $N_1(i, j)$ denotes the 1-ring neighbors of point (i, j) and $\|\cdot\|$ is the $\mathcal{L}_2$ norm. The smaller $S(i, j)$ is, the smoother the refractive vector field is at point (i, j) . Thus $S(i, j)$ can be used to estimate the significance of the noises in the extracted refractive vector field and the weight $\omega(i, j)$ should be calculated based on it. Therefore we define the weight for each vector by

$$\omega(i, j) = \begin{cases} 0, & \text{if } L(i, j) > \text{threshold} \\ \exp(-\alpha S(i, j)), & \text{if } L(i, j) \leq \text{threshold} \end{cases} \quad (\text{Eq. 3.7})$$

where α is a pre-defined constant and we set $\alpha = 0.1$ in all our experiments.

To find a reasonable $\mathcal{R}$ from the minimization problem, we let the solution space consist of bicubic B-spline surfaces. The mathematical equation of a parametric tensor product B-spline surface $\mathcal{R}(x, y)$ is

$$\mathcal{R}(x, y) = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} r_{i,j} N_i(x) N_j(y) \quad (\text{Eq. 3.8})$$

where $r_{i,j}$ are the B-spline coefficients to be determined; n_x, n_y are the numbers of the B-spline coefficients in x - and y -directions, respectively, which are initially chosen to be the maximum of 4 and one-tenth of the dimension of the region and then are adaptively adjusted during the fitting process according to the fitting error [32]; $N_i(x)$ and $N_j(y)$ are

normalized cubic B-spline functions [17]. Substituting (Eq. 3.8) into (Eq. 3.4), taking the partial derivatives of (Eq. 3.4) with respect to $r_{i,j}$, and letting them equal zero lead to a system of linear equations. The solution to the linear system gives the optimal B-spline surface, from which the target refraction vector field $\mathcal{R}$ can be recovered.

3.3.3 Experiments

We are now ready to summarize our transparent object matting and compositing method. First, we capture the images of the specially designed background at two different distances, with and without a transparent object in front of the background. Then we extract the refractive vector field as the matte for the transparent object. Next we apply piecewise B-spline fitting method to remove the artifacts of the refractive vector field. After that, we can perform compositing for any background at any distance away from the object. That is, given an arbitrary background at any distance, we can composite the transparent object into the background to synthesize a new image.

We compare our algorithm with the previous methods and also compare the compositing results with the ground truth. In our experiments, all the real captured or synthetic images are of the resolution of 512×512 . The comparisons are performed on both real captured transparent objects and a synthetic model. The synthetic transparent object is created using 3DS Max, whose material properties are set to be colorless and specularly refractive, which satisfy the assumptions and requirements of the real-time environment matting algorithm [24]. The object is then rendered in 3DS Max and the images generated are used as the inputs to our method or as the ground truth. To better illustrate the accuracy of the recovered refractive vector field, we choose two background images as the compositing background: the first one is a fruit background (Fig. 3.4.a) and the second one is a text background image with many letters (Fig. 3.4.b). Both background images contain much structure information, which helps to demonstrate the accuracy of

3.4.a: Fruit background

3.4.b: Text background

Figure 3.4: Two backgrounds used in the experiments.

the output refractive vector field. The more accurate refractive vector field we recover, the more structure details we can see in the area of the transparent object.

3.3.3.1 Accuracy

To verify the effectiveness and the accuracy of our proposed piecewise B-spline fitting method, we compare our method, the anisotropic filter method [24], the vector fitting method [35], and the ground truth. Fig. 3.5-3.8 are compositing examples of synthetic or real captured transparent objects with or without severe artifacts in the initially extracted environment matte data. In these figures, the first row shows the results of intermediate steps: the initially extracted refractive vector field, the segmentation result, and the weights in which the dark intensity indicates the severe noises or errors in the extracted environment matting data. The vector field is visualized by a color map in which the red and green channel values correspond to the x and y direction values after normalization. The second row shows the refined vector fields and the third row (or the fourth row) is for

the compositing results. The results in the first, second, third, and fourth columns are generated by the anisotropic filter method [24], the vector fitting method [35], and the proposed method, and the ground truth, respectively. In this experiments, the original matte data extraction time cost is about less than 1s for the 512×512 sample image, the piecewise fitting process takes about 2-3s to refine the matte data, and the final matte composition process is very fast, about 0.1s.

Fig. 3.5 shows the compositing results of a synthetic transparent glass with two different background images. Fig. 3.6 and 3.7 are examples of a real transparent object captured in two different orientations. These three examples all contain significant amount of noises and errors in the initially captured refractive vector field data, which can be seen from the weights or the initial vector fields. For the anisotropic filter method, we can see that the result vector fields are still full of artifacts, especially in the regions with severe errors. This is because the anisotropic filter method can only eliminate noises but not errors. For the vector fitting method, it can remove both noises and errors simultaneously, but it only identifies and rectifies errors in each horizontal line. Moreover, the vector fitting method fails when the axial symmetry assumption is not satisfied (see images in Fig. 3.7). Compared with these two methods, the proposed method can identify and eliminate artifacts globally and the compositing results are more accurate and clearer, even in error regions.

Fig. 3.8 shows another real captured example, in which the extracted environment matte data contain only a small number of noises. In this case, all the three methods can produce visually satisfactory results, but in general our method still produces a more accurate and smoother result which is very close to the ground truth, specially around the edge regions.

Figure 3.5: Compositing results of a synthetic transparent glass model with two different backgrounds.

Figure 3.6: Compositing results of a real captured transparent goblet whose extracted matte data contain severe artifacts.

Figure 3.7: Compositing results of the transparent goblet captured in a different orientation. The extracted environment matte data also contain severe artifacts.

Figure 3.8: Compositing results of a real captured transparent bowl, for which the extracted environment matte data are quite clean.

3.3.3.2 Comparison with the simple scaling method for distance-varying backgrounds

Now we consider composition with distance-changeable backgrounds, for which the state-of-the-art method needs to repeat the entire matte acquisition process for each new distance. In contrast, for our approach, the matte acquisition and the refractive vector field recovery process only need to be performed once. Based on the recovered refractive vector field, we can generate the compositing result for any background and any distance in a very convenient way. For comparison, we implement the simple scaling method, which scales the background to an appropriate size as the input for the method of [24]. The simple scaling method implicitly assumes that the foreground pixel lies on the refractive vector, which is generally not true as shown in Fig. 3.1. Therefore the simple scaling just gives an approximation of the refraction effect. Fig. 3.9 and 3.10 show the composition results with different distances between the transparent object and the background plane. To give more clearly comparison results, the composition result images are already cropped to focus on refraction effects of the transparent object. Compared with the ground truth rendered by 3DS Max, apparently our approach simulates the distance-changing effect better and produces more visually pleasing results, especially on the center sphere part of the object where the refracted background pattern information is changed drastically when the background distance changes. In this experiment, the computational cost is the same as the above experiment because we can easily calculate the new matte information for varying background distance.

3.3.4 Limitations

In our experiments, all objects are assumed to be colorless and specularly refractive and region blending effects is ignored. This is because the real-time environment matting method [24] that we used to extract matte data under two background distances adopts

Figure 3.9: Compositing results with a fruit background placed at different distances. In this example, as the background plane moving from near to far, the shape of the refracted grape shown in the center sphere part changes drastically, which are correctly simulated using our proposed algorithm while the previous scaling method fails to recover.

Figure 3.10: Compositing results with a text background placed at different distances. Similar to Fig. 3.9, as the background plane moves, the refracted text pattern seen through the transparent object changes a lot. Apparently, our proposed algorithm can give more accurate results than the previous method.

a simplified environment matte model based on the assumptions. Adopting more general methods to extract the initial environment matte data can certainly help remove such limitation but will bring in much more extra computational cost, which will be studied in the next chapter.

In addition, since our proposed method uses region segmentation and B-spline fitting scheme to recover the refractive vector field, it is assumed that the refractive vector field is piecewise smooth and the region segmentation result is roughly close to its real shape geometry. Note that the segmentation algorithm is processed on the extracted refractive vector field instead of the original sample image. In the case that some special transparent object has very complicated or much subtle geometry details, it is inevitable to get under-segmented regions and may miss some subtle refraction details. In this situation, our proposed method may not work very well to recover all shape details. Here we give an example for illustration, as shown in Fig. 3.11, the transparent dragon model has a lot of geometry variance on its body surface so that it is very hard to give proper region segmentation result using our current segmentation strategy¹. Thus, our proposed method produces a smoother composition result as shown in Fig. 3.11 while the ground truth contains a lot of fine refractions.

¹As far as we know, most of the existing automatic segmentation algorithms cannot handle such complicate case while only some interactive segmentation algorithms may give good results with certain amount of user interactions, which is beyond the scope of this thesis.

3.11.a: Segmentation Result

3.11.b: Our Result

3.11.c: Ground Truth

Figure 3.11: An example for which the proposed method does not work well. Here the dragon object contains very much subtle surface geometry that the segmentation method fails to recover, the fitting and composition result is smooth without too much fine refractions comparing to the ground truth.

Chapter 4

Compressive High Quality Environment Matting

In this chapter, we introduce our compressive sensing based high quality environment matting algorithm, which can significantly simplify the data acquisition process, reduce the number of sample images and the computational cost, suppress the measurement noise and faithfully recover the sparse environment matte data with the limited sample images.

4.1 Motivation and Our Approach

Generally speaking, the current existing environment matte extraction algorithms can be classified into two categories: real-time acquisition method and high quality extraction method. The real-time method [24] has very low complexity in matte acquisition process but may lack of accuracy because of possible measurement error and noise (although the accuracy limitation can be improved as discussed in Chapter 3, it is still only applicable to colorless and purely specular transparent objects). Therefore, to achieve a high accuracy environment matting result and handle more general circumstances, some existing high quality environment matting acquisition methods [24, 75, 121, 86, 143] are proposed which usually require the capturing of thousands of sample images and spend several

hours to capture sample images (see Table 4.1), which is very undesired for practical applications.

To reduce the complexity of data acquisition while still achieving high-quality environment matting, in this chapter, we incorporate the recently developed compressive sensing technology into the environment matting problem. In particular, the environment matting extraction problem is converted into a sparse light transport vector recovery problem, which can be well solved by compressive sensing techniques with a small number of measurements. Moreover, to have an efficient and robust solution, our compressive environment matting design takes into account the important properties of the light transport vector such as group clustering and Gaussian prior. We also propose a hierarchical method to reduce the computational cost. The data acquisition in our design becomes a simple process of capturing a 3-4 min video, which is much easier than the existing methods that capture a few thousand images.

We would like to point out that the environment matting problem can be considered as a special case of image based relighting and the compressive sensing theory has also been applied in the general image relight problem in [88, 100]. However, we argue that when applying compressive sensing in the specific problem of environment matting, there are some unique features. Our major contribution lies in the proposed solution to solve the compressive environment matting problem, which has some neat designs including hierarchical sampling, group clustering based recovery, Gaussian prior based denoising and video based data capturing that make the proposed high quality compressive environment matting framework more practical, efficient and effective.

4.2 Environment Matting Model and Properties

We begin by examining the environment matting equation introduced in [144], i.e.

$$C = F + (1 - \alpha)B + \sum_{i=1}^m \mathcal{R}_i \mathcal{M}(T_i, A_i) \quad (\text{Eq. 4.1})$$

where C , F and B are the colors of image, foreground, background respectively, α is the opacity value of the foreground object, $\mathcal{R}_i$ is the reflectance that describes the contribution of light emanating from environment to the object, and $\mathcal{M}$ is the “texture-mapping operator” calculating the average color of region A_i in background texture T_i . In general, this model of (Eq. 4.1) is comprehensive because it takes into account not only the foreground and background information but also the light refraction and reflection effects.

Considering the practical scenario of capturing sample images of transparent objects with one background monitor or projector and no other environment light source, F can be ignored since the camera only captures the background light that is not absorbed by the transparent object. Moreover, the second item and the third item in (Eq. 4.1) can be combined because they both describe the contributions from the background light source, although one is for direct transmission and the other is for refraction and reflection.

Based on the above analysis, similar to [86, 87], we model the environment matting and composition problem as

$$C = S + \rho \mathcal{T} \mathcal{L} \quad (\text{Eq. 4.2})$$

where ρ , called absorption index, is a scalar for one color channel describing the light absorption effects, $\mathcal{L}$ is an $N \times 1$ vector representing the N -dimensional background image and $\mathcal{T}$ is the light transport vector of $1 \times N$ illustrating the contribution of light emitted from each background region to an object pixel. Note that in (Eq. 4.2), we also introduce an offset term S , which is to compensate the camera response. This is because digital camera can only detect light intensity in a limited range. When the incident light intensity is small, the outgoing light may be too weak to be detected by digital camera, which will be treated as zero directly in the captured image.

In this way, the environment matting problem becomes: *given a set of background images $\mathcal{L}$ and the corresponding captured image colors C , how is the light transport vector*

$\mathcal{T}$ as well as ρ and S recovered? After obtaining all these parameters, we can easily simulate the light refraction and reflection effects under any new background image.

4.2.1 Property of light transport vector

Apparently, the essence of our environment matting model is to recover the light transport vector $\mathcal{T}$. Before introducing the recovery process, we summarize the important properties of the light transport vector as follows.

- **Value range:** Each element of $\mathcal{T}$ represents the percentage of contribution from a background pixel. Thus, elements of $\mathcal{T}$ should be non-negative and the sum of all the elements is equal to 1, i.e. $\|\mathcal{T}\|_1 = 1$.
- **Sparsity:** Typically, for a transparent object, there are only a small number of environment regions making contribution to a foreground pixel. That means the light transport vector is a sparse vector, where only elements for those limited background regions are non-zero while the others are all zeros.
- **Group clustering:** In addition to sparsity, another interesting phenomenon is that the regions that contribute to a certain foreground pixel are not randomly or uniformly distributed over the whole background plane. Actually most of them are neighboring and can be divided into several main parts (e.g. refraction part, reflection part and transmission part), which has been proven in [144, 24, 121]. This suggests that the non-zero elements $\mathcal{T}$ can be clustered into several local groups, which is a very useful prior for accelerating the CS-based recovery [59].
- **Gaussian prior:** More precisely, the light refraction and reflection effect of a transparent object should satisfy the BRDF property [24, 60, 119, 26, 29], which means that the non-zero elements of $\mathcal{T}$ should have 2D Gaussian distribution. This

is another very important constraint which is used later to verify the accuracy of our matte result.

To give a clear illustration of the light transport vector and its properties, we generate a composition result artificially and construct its corresponding light transport vector shown in Fig. 4.1. Particularly, the 256×256 contribution map in Fig. 4.1.a visualizes the contribution coefficient of each subregion in the background plane, where it is assumed that there are three separate regions (light refraction, reflection and transmission parts) contributing to the composition color result of one foreground pixel and the weights of each region are distributed according to the oriented Gaussian distributions [24]. Fig. 4.1.b shows the corresponding light transport vector $\mathcal{T}$ obtained by scanning the contribution map column by column. It can be seen that $\mathcal{T}$ is a sparse and non-negative vector with less than 5% non-zero elements clustered in several local groups and $\|\mathcal{T}\|_1 = 1$.

4.3 Solving Environment Matting Problem

Now we consider how to solve the environment matting problem. According to our model of (Eq. 4.2), for each foreground pixel, there are three variables (scalar S , ρ and vector $\mathcal{T}$) for each color channel and totally nine variables for RGB channels. It is hard to solve them all at one time. Thus, we choose a two-step approach to recover the unknowns. Once S , ρ and $\mathcal{T}$ are recovered, given any new background image, we can easily synthesize the resulting image of the transparent object under the new background.

4.3.1 Recovery of camera compensation and light absorption index

At the first step, we aim to recover camera compensation term S and the light absorption index ρ . Inspired by Zongker's methods, we project some constant image to the transparent object so as to eliminate the influence of light refraction and reflection. Since the

4.1.a: Contribution map

4.1.b: Corresponding light transport vector

Figure 4.1: Simulation of light refraction, reflection and transmission and the corresponding light transport vector. The background resolution is 256×256 and there are three separate parts contributing to certain foreground pixel with oriented Gaussian distribution for the weights of each region. The corresponding light transport vector is sparse with only 2806 non-zero elements out of 65536 elements.

Figure 4.2: Sample images of a transparent object in front of constant background images with different colors.

background plane has constant color, $\mathcal{L}$ can be written as the product of an intensity value b with an identity vector I . Further considering $\|\mathcal{T}\|_1 = 1$, we obtain $\mathcal{T} \cdot \mathcal{L} = b$. Thus, (Eq. 4.2) becomes

$$C = S + \rho b. \quad (\text{Eq. 4.3})$$

In this way, we can easily recover S and ρ for each color channel of every object pixel by capturing several sample images of the transparent object in front of constant background with different intensity values of b (see Fig. 4.2 for some examples).

4.3.2 Light transport vector recovery

Since $\mathcal{T}$ is a sparse vector, it can be faithfully reconstructed with a small number of measurements using compressive sensing algorithms. Consider that $\mathcal{T}$ is composed of the contribution coefficients of all the background regions to one foreground object pixel. By capturing a foreground pixel color result C under a given background pattern $\mathcal{L}$, we obtain one measurement of $\mathcal{T}$. If we construct an appropriate matrix $\mathcal{L}_{\mathcal{M}}$ with RIP, where each column is one background image pattern, to measure $\mathcal{T}$ many times and record all the resulted foreground pixel color C , $\mathcal{T}$ can then be recovered accurately through compressive sensing.

4.3.a: Coarse level

4.3.b: Fine level

Figure 4.3: Sample images of a transparent object in front of the random background patterns with different resolutions.

4.3.2.1 Background pattern generation

As pointed out in [12], only some special random matrices can be used as measurement matrix in compressive sensing. In our implementation, similar to [100], we choose the Bernoulli matrix, a binary matrix (containing either $+1$ or -1) drawn from a Bernoulli distribution. Considering a resolution of N for the background, we generate an $N \times M$ Bernoulli matrix $\mathcal{L}_{\mathcal{M}}$, where M is the number of measurement. For each column of $\mathcal{L}_{\mathcal{M}}$, a corresponding background pattern image is generated by mapping individual $+1$ and -1 elements to white and black colors, respectively. Fig. 4.3 shows the captured images of a transparent object in front of the random background patterns with different resolutions.

4.3.2.2 Recovery by compressive sensing

Under different background patterns, a series of sample images of the transparent object can be captured (see Fig. 4.3). For each foreground object pixel, we can rewrite (Eq. 4.2) into a matrix form:

$$\mathcal{C} = S \cdot I + \rho \mathcal{T} \mathcal{L}_{\mathcal{M}}, \quad (\text{Eq. 4.4})$$

where $\mathcal{C}$ now is a $1 \times M$ vector recording all the captured colors of the pixel under different background patterns, and I is a $1 \times M$ identity vector. Note that the $N \times M$ matrix $\mathcal{L}_{\mathcal{M}}$ is not the originally generated Bernoulli matrix, but the corresponding background image patterns, where each background image pattern is sampled to the same resolution as the captured images. Then, the target sparse vector $\mathcal{T}$ can be recovered by solving the following optimization according to the compressive sensing theory:

$$\min \|\mathcal{T}\|_0, \quad s.t. \quad \mathcal{T}\mathcal{L}_{\mathcal{M}} = \frac{\mathcal{C} - S \cdot I}{\rho}. \quad (\text{Eq. 4.5})$$

However, such a general solution is not cost-effective since it does not consider the properties of the light transport vectors.

Therefore, to make sure $\mathcal{T}$ is recovered faithfully and efficiently, during the recovery process we further add in the constraints based on the properties of $\mathcal{T}$ described in Section 4.2.1:

$$\min \|\mathcal{T}\|_0, \quad s.t. \quad \begin{cases} \mathcal{T}\mathcal{L}_{\mathcal{M}} = \frac{\mathcal{C} - S \cdot I}{\rho} \\ \forall i, t_i \geq 0 \\ \|\mathcal{T}\|_1 = 1 \\ \mathcal{T} \sim \mathcal{N}(c_x, c_y, \sigma_x, \sigma_y, \theta) \end{cases} \quad (\text{Eq. 4.6})$$

where t_i is the i -th element of $\mathcal{T}$, and $c_x, c_y, \sigma_x, \sigma_y, \theta$ are center, variance and rotation angle parameters of 2D Gaussian functions. Note that the last constraint in (Eq. 4.6) means that the non-zero elements of $\mathcal{T}$ should have 2D Gaussian distributions, which comes from the Gaussian prior and is important for noise removal. Most of the current compressive sensing algorithms deal with the problem of computing a sparse vector from “clean” measurement samples. However, for practical applications such as environment matting, the measurements often contain complex noise, which is hard to be identified or modelled. It is necessary to utilize the prior, i.e. the Gaussian prior here, to suppress the measurement noise by removing the non-zero elements in $\mathcal{T}$ that do not satisfy the Gaussian distribution property.

Moreover, the group clustering prior is directly used when solving (Eq. 4.6). Particularly, we treat all the non-zero elements as some local groups instead of several separate ones. As pointed out in [59], for a k -sparse vector $\mathcal{T} \in \mathbb{R}^n$, $\mathcal{O}(k \log(\frac{n}{k}))$ measurements are needed; but if k non-zero elements can be grouped into q groups, the number of required measurements can be reduced drastically to $\mathcal{O}(k + q \log(\frac{n}{q}))$. This is exactly the case in our problem, where q is typically no more than three based on our experiments. By using this group clustering prior, we are able to reduce the time complexity of the data acquisition and the recovery process significantly.

4.3.2.3 Hierarchical measurement and recovery scheme

As we know, for the environment matting problem, in order to generate a photo-realistic composition result, the resolution of the background pattern N should be sufficiently large, ideally the same as the resulting image resolution. However, this will lead to an extremely large dimension for the measurement matrix $\mathcal{L}_M$ and the light transport vector $\mathcal{T}$, which subsequently results in a very expensive computational cost to solve the constrained optimization problem (Eq. 4.6). Therefore, to accelerate the whole optimization process, we choose a hierarchical measurement and recovery scheme. The basic idea is to recover the light transport vector $\mathcal{T}$ progressively in a coarse-to-fine manner. Similar idea has also been exploited in [144, 86, 88]. In particular, as shown in Fig. 4.1.a, only small parts of the background area contributes to certain foreground pixel color while all the other parts are useless. Thus, an intuitive and efficient hierarchical scheme is designed as follows:

- (i) At the coarse level (see Fig. 4.3.a), we use compressive sensing to solve the constrained optimization problem (Eq. 4.5) with a very low-dimensional vector $\mathcal{T}$. The recovered $\mathcal{T}$ gives a rough estimation about where the contribution regions are, i.e. the rough location of those non-zero elements.

- (ii) At the fine level (see Fig. 4.3.b), we delete those useless dimensions in both $\mathcal{L}_{\mathcal{M}}$ and $\mathcal{T}$, increase the resolutions for the nonzero locations, and then resolve the optimization problem.

By using this hierarchical recovery method, the dimension and the computational cost of the optimization problem are reduced significantly.

4.4 Experiments

In this section, we conduct experiments to verify the effectiveness of the proposed compressive environment matting algorithm. In particular, we use Canon 5D Mark II digital camera to capture images of transparent objects and a 24" SAMSUNG LCD monitor to project background images. Unlike the measurement process in [86, 99], which generates adaptive background pattern in each iteration based on analyzing the previously captured sample images, our measurement process is non-adaptive, where the measurement matrix or the corresponding background patterns are generated and stored in advance. In this way, our data acquisition is simplified into a process of capturing a video of the transparent object in front of the monitor that projects various background patterns continuously. Note that during the data acquisition process, the positions of the camera, the transparent object and the monitor are fixed so as to make sure that the unknowns in (Eq. 4.4) remain unchanged.

4.4.1 Implementation of light transport vector recovery

As described in Section 4.3.2, we use compressive sensing with hierarchical structure to recover the light transport vector. In particular, for the monitor with a resolution of 1280×960 , we set the resolutions to be 16×12 and 128×96 for the coarse level and the fine level, respectively.

During the sparse vector recovery process, we adopt the dynamic group sparsity recovery algorithm (DGSR) [59] rather than the common ROMP method used in [88, 100] due to the following two reasons.

- (i) For a sparse vector with group clustering prior, the DGSR algorithm can recover it accurately with less computational cost (see Table 4.2) and less measurements [59].
- (ii) Most of the existing compressive sensing algorithms need to know the sparsity value of the vector before the recovery process starts. However, for our environment matting problem, it is impossible to know the exact sparsity value of the light transport vector $\mathcal{T}$ in advance. In addition, the sparsity values for distinct foreground points are usually different because of different shape, refraction and reflection properties. Fortunately, DGSR is not very sensitive to the input sparsity value and can recover the group clustering information dynamically without any prior.

Considering that a sparse vector can be steadily and accurately recovered if the number of measurements is above five times of the vector sparsity, we set the sparsity value of $\mathcal{T}$ to be 1/5 of the total measurements. Furthermore, to ensure the non-negative and sum constraints in (Eq. 4.6), we make some modifications to the original DGSR method. Specifically, in each iteration we solve the non-negative least square problem instead of the general least square problem. The iteration stops when the L_1 norm of the currently recovered vector is close to one.

4.4.2 Visual results of environment composition

We first verify the proposed algorithm using a real glass cup. In particular, we capture 24 images at coarse level (see 4.3.a) and 100 $\sim$ 400 sample images at fine level (see 4.3.b) for the designed hierarchical measurement. Then, the refractive light transport vector $\mathcal{T}$ of each object pixel is recovered using our proposed algorithm and subsequently composited

with a new background image. Note that since it is captured in real environment, the measurements inevitably contain some noise.

Fig. 4.4 shows the composition results of our methods with and without Gaussian prior. It can be seen that the results with Gaussian prior are always better than that without the prior. Moreover, with the decreasing number of image samples, the gap between the two methods becomes larger. At the case with only 100 sample images, the results of the method without Gaussian prior is full of artifacts. This is mainly because the number of measurements is insufficient to recover the sparse vector $\mathcal{T}$. In contrast, the method with Gaussian prior achieves much better performance since the additional Gaussian prior is able to effectively suppress the recovery errors in $\mathcal{T}$. Even with only 100 image samples, the result with Gaussian prior is still visually acceptable.

We also conduct some experiments in simulated environment. In particular, we use some well-known complex 3D model and obtain all the sample images through rendering the 3D model using 3DS Max. Such a scenario can be considered as an ideal case, since the capturing environment is well controlled and the measurement noise can be ignored. Fig. 4.5 shows the composition results of a glass elephant, where 150 sample images are used to recover 12288-dimensional sparse refractive light transport vectors. It can be seen that the composition result of the method with Gaussian prior is very close the ground truth. Its superior performance against that of the method without Gaussian prior is well demonstrated in the nose and foot regions. In these regions, the object shape changes more drastically, which results in the passing-through lights to be refracted and scattered stronger than other areas, makes the sparsity of $\mathcal{T}$ much bigger, and thus requires more sample images to recover the light transport vector $\mathcal{T}$. The additional Gaussian prior significantly reduces the number of required image samples and thus can perform better in these regions. Similar results can also be found in Fig. 4.6, where all these images are composited using the recovered matte data extracted using only 150 sample images within gaussian prior.

Figure 4.4: Composition results of our proposed methods with and without the Gaussian prior, where the former significantly outperforms the latter in terms of visual quality at the same number of sample images.

4.5.a: Ground truth

4.5.b: with Gaussian prior

4.5.c: w/o Gaussian prior

Figure 4.5: Composition results of a 3D glass elephant with 150 sample images. Note that all sample images and ground truth are rendered using 3DS Max.

4.4.3 More comparisons

To better illustrate the advantages of the proposed compressive environment matting framework, we compare our algorithm with the existing representative environment matting methods in Table 4.1. The information about the existing environment matting methods is directly copied or estimated from the corresponding publications.

From Table 4.1, we can see that the first two methods require less number of sample images but can only recover single-region or single-point mapping for each foreground pixel, which is certainly limited for high-quality environment matting and composition. Moreover, the real-time method [24] can only handle colorless and purely specular transparent objects. Among the other methods that can recover multiple-region mapping effects, our approach is the best one in terms of the number of required sample images and the data acquisition complexity. The scale of the number of sample images for our method only depends on the sparsity of the light transport vector, which is much smaller than the resolution of the background image. Note that the data acquisition process in our algorithms becomes a simple video capturing process that can be done in a few minutes while the other three methods take at least half an hour for data acquisition. Compared with our method without Gaussian prior, the one with the prior can reduce the

4.6.g: Refraction and reflection effects

4.6.h: Our composition result using 150 samples

4.6.i: Ground Truth

Figure 4.6: Our proposed composition result of several complex 3D models. From the results in the first column we can see that due to the complex shape, its refraction and reflection effects are very complicated, but we can still recover such complicated effects efficiently using only 150 sample images and generate very accurate composition results that close to the ground truth.

Table 4.1: Comparisons among different environment matting algorithms with a background image size of $n \times n$ and k denoting the sparsity of the light transport vector.

Methods	# of captured images when $n = 512$	Data acquisition time & method Matte extraction time	Feature
Zongker's method [144]	$\mathcal{O}(\log n)$ 18 images	5 min & image capture 10-20 min	Suitable object Support single-region mapping General transparent object
Real-Time method [24]	2 images	1 min & image capture 2 min	Support one-point mapping Colorless and pure specular object
Chuang's method [24]	$\mathcal{O}(n)$ 900 images	30 min & image capture 20 min	Support multiple-region mapping General transparent object
Wavelet method [86]	$\mathcal{O}(n)$ 2400 images	12 hours & image capture Not available	Decompose background into patterns General transparent object
Frequency method [143]	$\mathcal{O}(n)$ At least 2048 images	2 hours & image capture Not available	Support multiple-region mapping General transparent object
Our method w/o Gaussian prior	$\mathcal{O}(k \log k)$ 400 images	4 min & video capture Depending on object size	Support multiple-region mapping General transparent object
Our method with Gaussian prior	$\mathcal{O}(k \log k)$ 250-300 images	3 min & video capture Depending on object size	Support multiple-region mapping General transparent object

Table 4.2: Average time for the light transport vector recovery using the non-negative dynamic group sparsity recovery algorithm (NN-DGSR) with different settings.

	Hierarchical (seconds per 1000 pixels)	Direct
NN-DGSR	306 ± 5	538 ± 5

required image samples from 400 to about 250-300 and save at least 25% data acquisition time while producing visually accurate composition results.

In terms of the computational cost to extract the environment matte data, the major burden in our algorithm comes from the process to recover the light transport vector $\mathcal{T}$ in each color channel for every foreground point. Table 4.2 gives the average time taken to recover the light transport vector using the non-negative dynamic group sparsity recovery algorithm (DGSR) with different settings for the case in Fig. 4.4 with 400 sample images, where the direct method solves the constrained optimization problem with high dimension for once while the hierarchical one solves the optimization problem with low dimension for twice. The algorithms are running on an HP xw4400 workstation with Intel Core2 6700 CPU and 2G memory. Note that the time values given in Table 4.2 are the average results obtained by running the algorithms for many times over different parts of the transparent object. It can be seen that the hierarchical method is faster than the direct one and the matte extracting time is about 1 $\sim$ 2 hours in our experiments depending on the object size, which is acceptable for practical applications. Considering that the light transportation vectors of different foreground pixels are independent, we can use parallel processing such as GPU to accelerate the processing speed.

4.4.4 Comparison with compressive image relighting

Since environment matting can be considered as a special case of the image based relighting problem, in this section we give some discussions to compare our method with

the state-of-the-art compressive image relighting algorithm [88]. In general, both methods use compressive sensing technique to reduce the number of sample images and the hierarchical scheme to accelerate the measurement and recovery process.

However, there are some significant differences. First, we consider the unique properties of the environment matting problem such as the group clustering property, which might not exist in the general image relighting problem. With this group clustering property, we are able to reduce the number of image samples and accelerate both the data acquisition process and the matte data recovery process. In particular, [88] uses about 1000 non-adaptive illumination samples to recover a 128×128 sparse light transport vector while we only need about 250-300 samples to recover a 128×96 vector $\mathcal{T}$. Second, to fight with the common issue of measurement noise, [88] utilizes the similarity among neighboring pixels while we use the Gaussian prior to suppress the effects of the noise, which is more effective since it complies with the light BRDF property. Third, the hierarchical scheme in [88] is used to explore the similarity among neighboring pixels but the size of the measurement matrix was fixed all the time, while in our method the size of the measurement matrix in the finer level is much smaller since it is based on the results of the coarse level. Thus, the computational cost of our method is much less than that of [88].

Chapter 5

Joint Mesh Refinement and Light Function Recovery for High-quality 3D Reconstruction Under Unknown Illumination

In this chapter, we introduce a novel high quality shape reconstruction algorithm that integrates the lighting geometry property into conventional hybrid multi-view stereo and photometric stereo techniques to generate 3D object models with high-frequency surface details.

5.1 Motivation and Our Approach

As summarized in Chapter 2, MVS and PS are two common techniques for 3D reconstruction problem. MVS technique can generate rough 3D object models but miss high-frequency shape details, while PS technique is able to recover surface details but only under a fixed viewpoint. Joint MVS and PS techniques such as [64, 66, 62, 37, 125, 126, 137] can overcome the drawbacks of each individual technique at the costs of high computational complexity, large number of image samples, and the typical inflexible requirement of controlled illuminations.

In this chapter, we consider the problem of high-quality 3D reconstruction under

unknown general lighting conditions using the joint MVS and PS technique. We take into account the property of lighting geometry and propose to use total variation (TV) term to constrain the light function recovery. Similar to [126], our method uses MVS model as the initial input data. Then, we adopt the shading information to recover the missing high-frequency surface details to refine the initial MVS model. The main difference between our method and many previous methods is that both the illumination conditions and 3D vertex coordinates are treated as unknown variables to be optimized in our algorithm. Based on the Lambertian light reflectance model, we connect them together using the shading cues and optimize them simultaneously through a constrained energy minimization process.

The contribution of this research are twofold. First, we find that under many general illumination circumstances, surface illumination can be considered as piece-wise constant. Thus, we propose to apply total variation constraint on the light function. As far as we know, it is the first time that the total variation constraint is applied on illumination. Experimental results demonstrate the effectiveness of the proposed constraint. Second, in our framework, both the light function and the vertex positions are optimized simultaneously, while in the state-of-the-art method [126], they are recovered and optimized separately. Moreover, the method in [126] refines the mesh vertex by vertex due to the nonlinearity of the energy function while our method optimize the entire mesh together. Thus, our method, where the entire optimization process only costs a couple of minutes, has much lower computational complexity than [126].

5.2 Lighting geometry property

Before the description of the proposed algorithm in detail, some basic assumptions and the illumination model used in this research should be stated first. In particular, there are two basic assumptions:

- (i) The environment illumination is general, constant and far enough that can be modeled as directional lights;
- (ii) The target object is Lambertian object with a constant light reflectance parameter.

5.2.1 Lambertian lighting model

Since the environment illumination is distant enough, they can be approximated as a combination of many directional lights and here we use m different directional lights $\{l^k \in \mathbb{R}^3 | k = 1, \dots, m\}$ to represent the whole environment illumination. Then for an object surface point $v \in \mathbb{R}^3$, its intensity can be calculate as

$$c_v = \rho \times \sum_{k=1}^m \max(l^k \cdot n_v, 0) \times V(l^k, v) \quad (\text{Eq. 5.1})$$

where $V(l^k, v)$ is the binary visible function of v to directional light l^k and $\rho \in \mathbb{R}^1$ is the constant Lambertian reflectance coefficient. The above Lambertian reflectance model can be rewritten into the following form:

$$c_v = \sum_{V(l^k, v) > 0, l^k \cdot n_v > 0} \rho \times l^k \cdot n_v = l_v \cdot n_v \quad (\text{Eq. 5.2})$$

Here l_v is used to represent the combined lighting effects on v that determines the vertex intensity. Therefore, as shown in Fig. 5.1, instead of recovering the whole environment illumination $\{l^k | k = 1, \dots, m\}$, the visible function for each vertex to every possible incident directions $V(l^k, v)$ and then calculate $\sum_{V(l^k, v) > 0, l^k \cdot n_v > 0} \rho \times l^k$ for each surface vertex, which requires very high computational cost [126], in this thesis we directly focus on the variable l_{v_i} of each surface vertex v_i , which apparently satisfies the Lambertian reflectance model but with much less computational cost than [126].

5.2.2 Intensity constraints

According to (Eq. 5.2), the vertex intensity is determined by both surface normal n_v and the lighting function l_v . Therefore, once we obtain the vertex intensity c_v , its lighting

Figure 5.1: Illustration of lighting geometry on object surface. No matter what type of environment lights we have, for example, $\{l^k | k = 1, \dots, m\}$ from different directions, we can use l_v to represent the combined lighting effects on vertex v . Moreover, we can see that l_v would not change drastically between neighboring vertices, which means it has piece-wise constant property over the surface.

function and normal should satisfy the constraint:

$$c_v = l_v \cdot n_v \quad (\text{Eq. 5.3})$$

Moreover, enlighten by [126], for neighbor vertices, their intensity difference is also an important constraint that should be maintained:

$$c_{v_i} - c_{v_j} = l_{v_i} \cdot n_{v_i} - l_{v_j} \cdot n_{v_j} \quad (\text{Eq. 5.4})$$

where $v_j \in \mathcal{N}(v_i)$ is the one-ring neighbor vertex of v . Theoretically speaking, such intensity difference may be caused by both the normal change and light difference between points. But in most practical cases, it is due to the normal changes, that is, the high-frequency surface details that MVS technique fails to recover [126].

5.2.3 Piece-wise constant surface light

According to (Eq. 5.2), for a vertex v , its light condition l_v is determined by its normal n_v and visible function $V(l^k, v)$. Usually, such surface visible functions only change dramatically at edges and thus it is reasonable to assume that it is piece-wise constant over the surface. Therefore, the surface light condition l_v should also be piece-wise constant over surface (as shown in Fig. 5.1). To validate our hypothesis, we place the teapot model under several monochromatic environment lights¹ and render their results in 3Ds Max. From Fig. 5.2 we can see that the reflectance color is piece-wise constant over the surface. Therefore, to guarantee that such important lighting geometry property is satisfied during the optimization process, enlighten by [95, 16], we introduce the total variation constraint on vertex light function l_v in our framework. As far as we know, this is the first time that total variation constraint is used on illumination functions.

¹Since ρ is a constant scalar, the surface reflectance color just represents its lighting conditions if lights are monochromatic (e.g. red, green, blue).

Figure 5.2: Results of teapot model lit by several directional lights from different directions, from which we can see that its surface illumination condition is piece-wise constant. Here the lighting condition are converted to RGB color for illustration purpose.

5.3 Optimization Framework for Joint Mesh Refinement and Lighting Recovery

5.3.1 Problem Formulation

In this section we describe our new optimization framework that can accurately refine the object model and recover its light condition simultaneously. Similar to many previous methods, the reconstruction model of MVS techniques is used as the initial model, i.e., for each surface vertex we have the initial values of its position v^{in} .

Based on the discussions in Section 5.2, we include the following constraints in our framework:

- (i) **Fidelity constraint:** This term is to guarantee that the optimized vertex position v should be consistent with its initial MVS result:

$$E_f = \sum_{i=1}^N \|v_i - v_i^{in}\|^2$$

- (ii) **Photometric constraint:** This term aims to ensure that both the Lambertian reflectance model and the intensity difference between neighboring vertices are satisfied:

$$\begin{aligned} E_{ps} &= \sum_{i=1}^N \|l_{v_i} \cdot n_{v_i} - c_i\|^2 \\ &+ \sum_{e_{ij} \in E} \|(l_{v_i} \cdot n_{v_i} - l_{v_j} \cdot n_{v_j}) - (c_i - c_j)\|^2 \end{aligned}$$

where E is the set of all mesh edges.

- (iii) **Laplacian term:** This term can help avoid some singular or invalid triangles during the mesh optimization process. Also it can help make the mesh smooth.

$$E_{lap} = \sum_{i=1}^N \|v_i - \bar{v}_i\|^2$$

where $\bar{v}_i$ is the average of one-ring neighbors of v_i .

- (iv) **TV light constraint:** Since the surface light function is piece-wise constant, we propose to use the TV term on l_{v_i} to ensure such property during the optimization process:

$$R_{tv}(l) = \sum_{i=1}^N \|\nabla l_{v_i}\|$$

Finally, we formulate the 3D reconstruction problem as an energy optimization framework:

$$\min\{\frac{\alpha}{2}E_f + \frac{\beta}{2}E_{ps} + \frac{\eta}{2}E_{lap} + R_{tv}(l)\} \quad (\text{Eq. 5.5})$$

where α , β and η are weights for different terms.

5.3.2 Proposed Optimization Solver

It is difficult to calculate the optimal solution of our proposed optimization problem (Eq. 5.5) directly because of the TV term. Enlighten by [130], we propose to use a modified Augmented Lagrangian Method (ALM) to solve (Eq. 5.5).

The basic idea of ALM method is to introduce a temporary variable to substitute the gradient term in the original problem and use Lagrangian method to constraint its fidelity. In particular, our problem (Eq. 5.5) can be rewritten into the following form:

$$\begin{aligned} \mathcal{G}(v, l, p; \lambda) &= \frac{\alpha}{2}E_f + \frac{\beta}{2}E_{ps} + \frac{\eta}{2}E_{lap} + \lambda \cdot (p - \nabla l) \\ &+ \frac{r}{2}\|p - \nabla l\|^2 + R_{tv}(p) \end{aligned} \quad (\text{Eq. 5.9})$$

Algorithm 1 Modified Augmented Lagrangian Method

Input: Initial MVS mesh $\{v_i^{in} \in \mathbb{R}^3 | i = 1, \dots, N\}$, Vertex intensity $\{c_i \in \mathbb{R}^1 | i = 1, \dots, N\}$ and r, ϵ ;

Output: Optimal $\{v_i \in \mathbb{R}^3 | i = 1, \dots, N\}$ and vertex light $\{l_{v_i} \in \mathbb{R}^3 | i = 1, \dots, N\}$;

1: Initialization: $v^0 = v_i^{in}$, $l^0 = 0$, $p^0 = 0$, $\lambda^0 = 0$;

2: **repeat**

3: Solve u -sub problem: for a given p , solve l and v :

$$\min \left\{ \frac{\alpha}{2} E_f + \frac{\beta}{2} E_{ps} + \frac{\eta}{2} E_{lap} + \frac{r}{2} \left\| \left(p^k + \frac{\lambda^k}{r} \right) - \nabla l \right\|^2 \right\} \quad (\text{Eq. 5.6})$$

4: Solve p -sub problem: for a given v and l , solve p :

$$\min \{ R_{tv}(p) + \lambda^k \cdot p + \frac{r}{2} \| p - \nabla l^{k+1} \|^2 \} \quad (\text{Eq. 5.7})$$

5: Update Lagrange multiplier λ :

$$\lambda^{k+1} = \lambda^k + r(p^{k+1} - \nabla l^{k+1}) \quad (\text{Eq. 5.8})$$

6: **until** $\sum_{i=1}^N \| l_{v_i}^{k+1} - l_{v_i}^k \|^2 < \epsilon$

where p is the variable that is used to substitute ∇l , λ is Lagrange multiplier and r is a positive constant as the step size of iteration. According to [130], (Eq. 5.9) can be solved by dividing into three sub-problems that to be optimized iteratively as illustrated in Algorithm 1.

Note that in our modified augmented lagrangian method, the most important step is the u -sub problem, which can refine both the vertex position and surface light function simultaneously. Because of the normalization process during the calculations of both surface normal n_v and the gradient coefficients on triangular mesh, it is apparently a non-convex problem to the vertex position v and light l . Therefore, we modify the original u -sub problem in [130] and proposed to solve its equivalent problem (Eq. 5.6) instead, which is a non-linear least square optimization problem that can be solved through Levenberg-Marquardt algorithm [74] in our framework.

Figure 5.3: Illustration of vertex normal n_a . For a triangular mesh, the normal n_a for vertex v_a is actually calculated according to its one-ring neighboring points.

5.3.3 Jacobian matrix construction

In Levenberg-Marquardt algorithm, its essential process is to construct the Jacobian matrix which records the partial derivative of the optimization function to target variables separately, including the partial derivative of normalized surface normal to vertex XYZ coordinates and the partial derivative of the gradient of surface light to both vertex XYZ coordinates and light variables. So in this section we will introduce some details about how to calculate the above two partial derivatives.

For triangular meshes, the vertex normal is calculated as the sum of its neighboring triangle faces. That is, as shown in Fig. 5.3, for vertex v_a , its one-ring neighbors are

$\{v_b, v_c, v_d, v_e, v_f, v_g\}$, then its normal N_a is calculated as:

$$N_a = N_{f_{abc}} + N_{f_{acd}} + N_{f_{ade}} + N_{f_{aef}} + N_{f_{afg}} + N_{f_{agb}}$$

where $N_{f_{abc}} = (v_b - v_a) * (v_c - v_a)$ and $*$ represents the cross product operation. Summing all terms together, N_a actually equals to:

$$\begin{aligned} N_a = & v_b * v_c + v_c * v_d + v_d * v_e \\ & + v_e * v_f + v_f * v_g + v_g * v_b \end{aligned} \quad (\text{Eq. 5.10})$$

n_a is the normalized result of N_a , that means, $n_a = \frac{N_a}{\|N_a\|}$. Therefore, we can calculate the derivative of n_a to anyone of its neighbor vertices, for example, v_c , as:

$$\begin{aligned} \frac{\partial n_a}{\partial v_c} = & \frac{Jacobian(N(v_c), v_c)}{\|N_a\|} \\ - & \frac{transpose(N_a) * N_a * Jacobian(N(v_c), v_c)}{\|N_a\|^3} \end{aligned}$$

where $N(v_c) = v_b * v_c + v_c * v_d$. Note that when we calculate $\frac{\partial n_a}{\partial v_c}$, only v_c is variable while all other vertex positions and the vertex normal N_a are all treated as constant vectors. Similarly, the derivative of n_a to others neighboring vertices can be obtained in the same way. For more details of $\frac{\partial \nabla l}{\partial v}$ and $\frac{\partial \nabla l}{\partial t}$, please refer to [127, 128].

In another aspect, since all such partial derivatives are calculated using the target vertex and its one-ring neighbor vertices, such jacobian matrix is very sparse actually.

5.4 Experiments

To verify the efficiency and effectiveness of our framework, we validate our algorithm both on the synthesized bunny example, the real world fish example used in [126] and the temple and dino dataset from the Middlebury benchmark [4], which we do not have any prior knowledge about its environment light conditions. For the bunny example, we place the original bunny model under several directional lights and render its surface

intensity accordingly. Then we use a smoothed bunny as the input of our framework and refine the mesh as well as calculate its surface light function using our proposed algorithm. For the fish, temple and dino examples, in our framework, firstly we use the conventional MVS algorithm to calculate an initial smooth 3D model as the input of our framework. Specifically, our initial models are generated using the volumetric reconstruction methods [107], which has a very high rank and score for its reconstruction accuracy and computational error rate in the benchmark [4]. To show that our algorithm is not sensitive to the initial MVS models, we purposely smooth the initial reconstruction results and use the smoothed models as the input to our optimization framework. Next, considering that the camera parameters are already known [4] or can be recovered during the MVS process, we map the input MVS model to the real captured sample images (as shown in Fig. 5.4) in different views and compute its average intensity result as c_i for each surface vertex v_i . Finally the target object model and its surface illumination conditions are refined and recovered as specified in Algorithm 1.

5.4.1 Visual results

In this section, we show the results of mesh refinement, especially focusing on the accuracy of surface details. Firstly for the synthesized bunny example, the first two columns in Fig. 5.5 show a very smooth bunny model without much surface details and its surface intensity as the input of our framework. By utilizing the surface intensity constraints and surface lighting prior, our proposed can generate a more accurate bunny mesh and recover its corresponding piece-wise surface lighting functions, which are both very close to the ground truth. For this bunny example, we choose the parameters that $\alpha = 10.e + 8, \beta = 1.0e + 4, \eta = 10000, r = 0.5, \epsilon = 1.0e - 10$. Next, for the fish dataset, Fig. 5.6 shows the comparison between our result and that of Wu’s method [126], which represents the state-of-the-art approach on utilizing lighting and shading information for

Figure 5.4: Examples of real captured sample images of temple and dino objects.

3D reconstruction under general unknown illumination. The result of Wu’s method is kindly provided by the authors, which has 200k vertices while our recovered mesh model has only 40k vertices. Although our reconstructed fish model has much less vertices, the visual quality of our result is comparable to that of Wu’s method.

Moreover, for the next temple and dino dataset, since there is no result available for Wu’s method for this dataset, we compare our result with the volumetric MVS algorithm [107] which is one of the high-performance MVS methods. For the temple example, which is a challenging one since the object is of complex shape and might not meet the Lambertian surface assumption well. Even for such a case, our method using 48000 vertices can still generate more details than [107]. As shown in Fig. 5.7, the first column shows the initial smooth temple model as the input of our proposed framework while the second column shows the mesh vertex intensity information extracted from real captured images (as shown in Fig. 5.4); the reconstructed temple model using conventional MVS method [107] are shown in the third column which gives us an roughly accurate result but fails to recover many carved surface details, especially on its top parts and four pillars. On the contrary, our proposed mesh optimization framework can generate high quality 3D temple model with better surface details as shown in the last column of Fig. 5.7, especially in its top region and four pillar parts (The close view for the surface details comparison are shown in Fig. 5.8). For this temple example, we choose the parameters that $\alpha = 10.e + 6, \beta = 1.0e + 5, \eta = 100, r = 0.5, \epsilon = 1.0e - 10$. Also, Fig. 5.9 shows the comparison results of the dino model, which has very smooth body and uneven back squama. The MVS technique [107] generates a model with uneven surface and missing back squama details as shown in the first row of Fig. 5.9. Our algorithm can refine such uneven model, make its body smooth and recover back squama surface details. Moreover, our proposed algorithm can even recover the missing eye and mouth structure as shown in Fig. 5.10. For this dino example, we set the parameters

$\alpha = 1.0e + 10, \beta = 1.0e + 5, \eta = 1.0e + 8, r = 0.5, \epsilon = 1.0e - 10$. Note that here we choose a large η to make the dino surface smooth after mesh refinement.

Figure 5.5: Results of the synthesized bunny example. For this bunny example, the input of our framework is a smooth bunny model (shown in the first column) and its rendered images (shown in the second column). Then we can refine the mesh and recover its piece-wise surface lighting function as shown in the third and fifth column, respectively. Compared with the ground truth (column 4 and 6), we can see that our proposed algorithm can generate high-quality mesh with high frequency shape details and recover the lighting function close to the ground truth.

In another aspect, Fig. 5.11 shows the recovered surface light result for both the temple and dino examples, where we can see that our recovered light results satisfy the piece-wise constant constraint.

Figure 5.6: Comparison with Wu's algorithm [126]. For the mesh, our algorithm can refine the initial smooth model and recover mesh surface details similar to [126] but using a more reasonable optimization framework and at a very low computational cost.

5.4.2 Computational cost

In this section we discuss the computational cost of our algorithm in terms of time cost and memory cost.

In our experiment, we validate our algorithm on the middlebury multi-view stereo benchmark dataset [4] where the initial MVS models can be directly obtained and loaded into our framework. Then, as illustrated in Algorithm 1, the object model and its surface light are optimized iteratively until convergence. In each iteration, its computational time cost is mainly on the u -sub problem (Eq. 5.6) while the p -sub problem (Eq. 5.7) and the λ update process (Eq. 5.8) can be solved at a much faster speed [129, 130]. For the u -sub problem (Eq. 5.6), the Levenberg-Marquardt algorithm [74] is adopted in our framework, which takes about 8 minutes to solve (Eq. 5.6) for both temple and dino examples. The whole iteration will be repeated about 3-4 times until the stopping criteria are satisfied, which means our proposed energy minimization problem is solved in about half an hour. Thus, the computational time cost of our proposed algorithm is much less than [126], which takes half an hour to recover the environment lighting and the visibility function for each vertex, and takes about 2 hours to refine an object mesh. The time saving is because of three reasons: 1) for the environment light conditions, we focus on the combined light function l_v for each vertex directly while the method in [126] chooses to calculate the whole environment lights and the visible function for each vertex explicitly,

Figure 5.7: Result comparison for the temple model. Here the input smooth model of our proposed algorithm are shown in the first column; the second column shows the vertex intensity of the model which are extracted from real captured images as shown in Fig. 5.4; the reconstructed temple model using conventional MVS algorithm [107] and our proposed algorithm are shown in the last two columns respectively.

Figure 5.8: Close view of the pillar region of Temple. It can be seen that our proposed algorithm can generate more carved pillar details than [107] as reflected by the photometric cues on the object surface.

Figure 5.9: Result comparison for the Dino model. The first row shows the reconstructed Dino model using conventional MVS algorithm [107], the second row shows the obtained surface intensity value for each vertex while the last row shows the results of our method.

Figure 5.10: Comparison of dino head details. Our proposed algorithm can recover the eye and mouth structure for the dino while the conventional MVS algorithm [107] fails.

which is very time-consuming; 2) our approach refines object mesh and calculates the surface lighting conditions simultaneously while the method in [126] does it separately; 3) we propose to solve a non-linear least square problem for mesh optimization so that we can refine the positions for all vertices together in each step while the method in [126] refines the object mesh vertex by vertex in a fixed order.

One the other hand, since we optimize both mesh vertex coordinates and light information at the same time, the memory cost of our framework is large. Similar to the time cost, in our algorithm, the main memory cost also lies on the u -sub problem because we need to first calculate the corresponding Jacobian matrix J and residual vector f for (Eq. 5.6) and then solve a large linear equation $(J^T J - \mu I)x = J^T f$ to calculate the incremental value for both v and l_v . To be precisely, suppose the object mesh has N vertices, E edges and F faces, then we need $N \times 3 + N \times 3$ rows to calculate the derivative of E_f term, $N + E$ rows for E_{ps} term, $N \times 3$ rows for E_{lap} term and $F \times 9$ for the last term in (Eq. 5.6), which means that the size of the corresponding Jacobian matrix J is $(N \times 10 + E + F \times 9) \times (N \times 6)$ and the size of the linear problem is $(N \times 6) \times (N \times 6)$ in each iteration step. Taking the temple object as an example, which has about 48000

Figure 5.11: The calculated surface illumination results for both temple and dino dataset. The surface illumination results are converted into RGB color for illustration purpose. We can see that the illumination results satisfy the piece-wise constant constraint.

vertices and 96000 faces, we need about 10G memory for the whole optimization process without too much code optimization.

5.4.3 Convergence

In this section we discuss the convergence issue of the proposed optimization algorithm. The essential part of our framework is to solve the complex energy optimization problem (Eq. 5.5) to refine the mesh and calculate the vertex light property simultaneously, which contains both the total variation constraint on vertex lighting function and non-linear constraints on vertex positions. In the proposed optimization algorithm, the modified ALM algorithm is used to separate the original complicated problem into some sub-problems and optimize them iteratively. (Eq. 5.6) is the most critical one among all sub-problems. It can be considered as the combination of the following two problems: 1) Fixing mesh vertex position, optimize the lighting functions that satisfy total variation constraint; 2) Fixing the environment lighting functions, optimize the vertex positions according to Lambertian lighting model. Apparently the former problem is convex while the latter one is not. Thus, our algorithm cannot guarantee to achieve the global optimal solution theoretically.

On the other hand, our algorithm is based on the input of a smooth MVS reconstructed mesh, which is already close to the target roughly. The proposed optimization is mainly meant to further refine the initial mesh to generate more surface details. Thus, even a local optimum will still lead to better reconstruction performance. Our various experiments, both on the synthesized bunny and the real world fish, temple and dino dataset, demonstrate that such local optimal result is acceptable for practical applications. Some recent work [126] also shows that local minima can give improved results.

5.4.4 Limitations

To be noted that our proposed algorithm still has some limitations. For one thing, the memory cost of our algorithm is very high because both the vertex position and its light condition are variables, so that currently we can not handle extremely large mesh with too many vertices. For another, similar to [126], the purpose of our algorithm is to refine the existing mesh model rather than reconstruction so that we require an roughly accurate input object mesh.

Chapter 6

Conclusion and Future Works

6.1 Conclusion

In this thesis, we have investigated the fundamentals of the two specific lighting geometry related problems: environment matting and 3D reconstruction, and reviewed the existing solutions for these two problems. We have proposed a novel flexible environment matting and compositing algorithm, a new compressive environment matting method and a novel joint mesh refinement and light function recovery framework to solve the above two problems effectively and efficiently.

To improve the flexibility of conventional environment matting and compositing technique, we have described a new method for flexible and accurate transparent object matting and compositing. We first introduce the new concept of refractive light vector and propose to use a refractive light vector field as the new representation for environment matte, which allows us to easily composite a transparent object into an arbitrary new background placed at any distance. Such flexibility cannot be provided by the existing environment matting algorithms, which require to repeat the whole complicated and time-consuming matting extraction process for any new distance. Moreover, to provide accurate environment matte data, we have also proposed a piecewise vector field fitting algorithm using bicubic B-splines, which can eliminate both noises and errors efficiently.

Compared with the previous methods, our approach can produce more robust and accurate compositing results.

To reduce the computational cost of high-quality environment matting extraction problem, we have proposed a novel compressive environment matting framework that incorporates the recently developed compressive sensing theory into the environment matting problem. Particularly, based on our analysis and validation, our proposed method incorporates the important properties of the light transport vector such as group clustering and Gaussian priors, with a hierarchical sampling and sparse vector recovery scheme. The experiment results show that compared with the existing high-quality environment matting methods, our proposed algorithm can recover the refraction and reflection properties of transparent objects accurately with much simplified data acquisition process, much less computational time cost and much less number of sample images.

To improve the accuracy of 3D reconstruction problem, we have developed a new high-quality 3D reconstruction algorithm that integrates the lighting geometry property into conventional hybrid multi-view stereo and photometric stereo techniques to generate 3D object models with high-frequency surface details. In our framework, both the mesh vertex position and the corresponding illumination condition are unknown variables that will be optimized in our research. Based on our analysis about the lighting conditions, its piece-wise constant property is proposed so that we propose to use the total variation constraint on it. To the best of our knowledge, it is the first time that total variation constraint is used on lighting problem. According to the surface shading cues, we can refine the object models and calculate the lighting conditions simultaneously. The experiment results show that our approach can generate the 3D models with more high-frequency surface details than conventional MVS algorithms. Compared with the previous methods, our propose method is reasonable, can provide a significant computational saving and is compatible with traditional MVS methods without capturing any extra images.

6.2 Future Directions

Firstly, note that in chapter 3, to simplify the data acquisition process we use the real-time environment matting method [24] to extract matte data under two background distances, which assumes that the target transparent object is colorless and specular refractive. Therefore it would be interesting to extend our novel refractive light vector field representation to handle more general situations that include common transparent objects or other objects with various kinds of light interaction effects. Moreover, consider that in our proposed flexible environment matting algorithm, the position of camera and object are fixed while only the background image plane is changeable. So it would be worthwhile to construct a more general framework where both the background plane and camera position can be changed arbitrarily. In addition, most environment matting algorithms judge the quality of the compositing results simply based on how the results visually look like the ground truth. When the ground truth (real captured image or rendered image) is available, existing objective metrics such as MSE and PSNR might be used to evaluate the matting and compositing performance. However, it is observed that the recovered structure information in the compositing results is usually crucial to the appearance, which is not well measured by those existing metrics. How to properly evaluate various environment matting and compositing methods warrants future investigation.

Secondly, since the environment matting and compositing problem is the inverse problem of image based relighting, it would be interesting to introduce our group clustering and oriented gaussian priors or incorporate other kinds of light constraints into general image based relighting problem to simplify the data acquisition process, reduce the computational cost and improve the accuracy of existing image based relighting algorithms.

Next, consider the fact that the Lambertian reflectance model and vertex intensity are adopted in our research, it is obvious that the proposed joint mesh refinement and light function recovery framework can be extended in the following directions. For one

thing, in many real situations, the Lambertian reflectance model may not be suitable or satisfied, e.g. for specular, translucent and transparent materials, so it would be worthwhile to modify our algorithm to handle other kinds of lighting models. For another, we can further investigate to use other kinds of illumination effects, such as, highlights, specular, light scattering, etc. as the photometric constraint for different specifications of 3D reconstruction problems. Last but not least, since our algorithm is compatible with conventional MVS algorithm without capturing any extra images, we can also extend our algorithm to handle dynamic objects or scenes to improve the reconstruction accuracy.

Last but not least, while the 3D reconstruction problem of opaque objects has been well studied, how to reconstruct transparent objects is still a big challenge due to the light reflection, refraction and specular effects. In another aspect, such complex lighting effects can be well recorded through environment matting and compositing technique. Therefore, it is of great interest to develop a novel algorithm that combines the 3D reconstruction technique and environment matting technique together to generate the 3D models of transparent objects.

References

- [1] Beam tracing algorithm. http://en.wikipedia.org/wiki/Beam_tracing.
- [2] Bidirectional reflectance distribution function. http://en.wikipedia.org/wiki/Bidirectional_reflectance_distribution_function.
- [3] Lambertian reflectance model. http://en.wikipedia.org/wiki/Lambertian_reflectance.
- [4] Middlebury multi-view stereo evaluation. <http://vision.middlebury.edu/mview/eval/>.
- [5] Phong reflection model. http://en.wikipedia.org/wiki/Phong_reflection_model.
- [6] Photon mapping algorithm. http://en.wikipedia.org/wiki/Photon_mapping.
- [7] Precomputed radiance transfer algorithm. http://en.wikipedia.org/wiki/Precomputed_Radiance_Transfer.
- [8] Ray tracing algorithm. [http://en.wikipedia.org/wiki/Ray_tracing_\(graphics\)](http://en.wikipedia.org/wiki/Ray_tracing_(graphics)).
- [9] Sameer Agarwal, Satya P. Mallick, David J. Kriegman, and Serge Belongie. On refractive optical flow. In *ECCV (2)*, pages 483–494, 2004.
- [10] Neil G. Alldrin, Todd Zickler, and David J. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA*, 2008.

REFERENCES

- [11] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1:205–220, April 1992.
- [12] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation (Jan.)*, 28(3):253–263, 2008.
- [13] Ronen Basri, David W. Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.
- [14] Moshe Ben-Ezra and Shree K. Nayar. What does motion reveal about transparency? In *ICCV*, pages 1025–1032, 2003.
- [15] James F. Blinn and Martin E. Newell. Texture and reflection in computer generated images. *Commun. ACM*, 19(10):542–547, 1976.
- [16] Peter Blomgren and Tony F. Chan. Color tv: total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.
- [17] C. De Boor. On calculating with b-splines. *J. Approx. Theory*, 6:50–62, 1972.
- [18] M. Born and E. Wolf. *Principles of Optics*. October 1999.
- [19] Derek Bradley, Tamy Boubekeur, and Wolfgang Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA*, 2008.
- [20] Neill D. F. Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part I*, pages 766–779, 2008.

- [21] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
- [22] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.
- [23] Biswarup Choudhury and Sharat Chandran. A survey of image-based relighting techniques. In *GRAPP 2006: Proceedings of the First International Conference on Computer Graphics Theory and Applications, Setúbal, Portugal, February 25-28, 2006*, pages 176–183. INSTICC - Institute for Systems and Technologies of Information, Control and Communication, 2006.
- [24] Y-Y. Chuang, D.E. Zongker, J. Hindorff, B. Curless, D.H. Salesin, and R. Szeliski. Environment matting extensions: towards higher accuracy and real-time capture. In *SIGGRAPH '00*, pages 121–130, New York, NY, USA, 2000. ACM Press.
- [25] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.
- [26] Qiang Dai, Jiaping Wang, Yiming Liu, John Snyder, Enhua Wu, and Baining Guo. The dual-microfacet model for capturing thin transparent slabs. *Computer Graphics Forum*, 28(7):1917–1925, 2009.
- [27] Kristin J. Dana, Shree K. Nayar, Bram Van Ginneken, and Jan J. Koenderink. Reflectance and texture of real-world surfaces authors. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 151–, Washington, DC, USA, 1997. IEEE Computer Society.
- [28] Andreas Dante and Mike Brookes. Precise real-time outlier removal from motion vector fields for 3D reconstruction. In *ICIP (1)*, pages 393–396, 2003.
- [29] Charles De Rousiers, Adrien Bousseau, Kartic Subr, Nicolas Holzschuch, and Ravi Ramamoorthi. Real-time rough refraction. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. ACM, 2011.

- [30] Amaël Delaunoy, Emmanuel Prados, Pau Gargallo, Jean-Philippe Pons, and Peter F. Sturm. Minimizing the multi-view stereo reprojection error for triangular surface meshes. In *Proceedings of the British Machine Vision Conference 2008, Leeds, September 2008*, 2008.
- [31] Ronald A. DeVore, Björn D. Jawerth, and Bradley J. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on Information Theory*, 38(2):719–746, 1992.
- [32] P. Dierckx. An algorithm for surface fitting with spline function. *IMA J. Numer. Anal.*, 1:267–283, 1981.
- [33] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [34] Mark S. Drew, Yacov Hel-Or, Tom Malzbender, and Nasim Hajari. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image Vision Computing*, 30(4-5):317–331, May 2012.
- [35] Qi Duan, Jianfei Cai, and Jianmin Zheng. Vector field fitting for real-time environment matting of transparent objects. In *Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt*, pages 3229–3232. IEEE, 2009.
- [36] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, Ting Sun, K.F. Kelly, and R.G. Baraniuk. Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):83–91, march 2008.
- [37] Carlos Hernández Esteban, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(3):548–554, 2008.
- [38] Pascal Fua and Yvan G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1):35–56, 1995.

- [39] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(8):1362–1376, 2010.
- [40] Pau Gargallo, Emmanuel Prados, and Peter F. Sturm. Minimizing the reprojection error in surface reconstruction from images. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–8, 2007.
- [41] Andrew S. Glassner. An overview of ray tracing. pages 1–31, 1989.
- [42] Michael Goesele, Brian Curless, and Steven M. Seitz. Multi-view stereo revisited. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, pages 2402–2409, 2006.
- [43] Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, and Steven M. Seitz. Multi-view stereo for community photo collections. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–8, 2007.
- [44] Dan B. Goldman, Brian Curless, Aaron Hertzmann, and Steven M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(6):1060–1071, 2010.
- [45] V.K. Goyal, A.K. Fletcher, and S. Rangan. Compressive sampling and lossy compression. *IEEE Signal Processing Magazine*, 25(2):48–56, march 2008.
- [46] Ned Greene. Environment mapping and other applications of world projections. *IEEE Comput. Graph. Appl.*, 6(11):21–29, 1986.
- [47] Jinwei Gu, Shree Nayar, Eitan Grinspun, Peter Belhumeur, and Ravi Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 845–858, Berlin, Heidelberg, 2008. Springer-Verlag.

REFERENCES

- [48] Martin Habbecke and Leif Kobbelt. A surface-growing approach to multi-view stereo reconstruction. In *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), 18-23 June 2007, Minneapolis, Minnesota, USA, 2007*.
- [49] Ziyad S. Hakura, John M. Snyder, and Jerome E. Lengyel. Parameterized environment maps. In *I3D '01: Proceedings of the 2001 symposium on Interactive 3D graphics*, pages 203–208, New York, NY, USA, 2001. ACM.
- [50] J.P. Haldar, D. Hernando, and Zhi-Pei Liang. Compressed-sensing mri with random encoding. *IEEE Transactions on Medical Imaging*, 30(4):893–903, april 2011.
- [51] J G Hayes and J Halliday. The least-squares fitting of cubic spline surface to general data sets. *J. Inst Math. Appl*, 14:89–103, 1974.
- [52] David J. Heeger and James R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238, New York, NY, USA, 1995. ACM.
- [53] Carlos Hernández, George Vogiatzis, Gabriel J. Brostow, Björn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–8, 2007.
- [54] Aaron Hertzmann and Steven M. Seitz. Shape and materials by example: a photometric stereo approach. In *Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition*, CVPR'03, pages 533–540, Washington, DC, USA, 2003. IEEE Computer Society.
- [55] Aaron Hertzmann and Steven M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(8):1254–1264, 2005.
- [56] Vu Hoang Hiep, Renaud Keriven, Patrick Labatut, and Jean-Philippe Pons. Towards high-resolution large-scale multi-view stereo. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pages 1430–1437, 2009.

- [57] Eli Horn and Nahum Kiryati. Toward optimal structured light patterns. In *International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97), May 12-15, 1997, Ottawa, Ontario, Canada*, pages 28–35. IEEE Computer Society, 1997.
- [58] Jun-Wei Hsieh, Wen-Fong Hu, Chia-Jung Chang, and Yung-Sheng Chen. Shadow elimination for effective moving object detection by gaussian shadow modeling. *Image Vision Computing*, 21(6):505–516, 2003.
- [59] Junzhou Huang, Xiaolei Huang, and Dimitris N. Metaxas. Learning with dynamic group sparsity. In *IEEE 12th International Conference on Computer Vision, ICCV 2009*, pages 64–71, Kyoto, Japan, 2009.
- [60] Ivo Ihrke, Gernot Ziegler, Art Tevs, Christian Theobalt, Marcus A. Magnor, and Hans-Peter Seidel. Eikonal rendering: efficient light transport in refractive objects. *ACM Transactions on Graphics*, 26(3):59, 2007.
- [61] Michal Irani, Benny Rousso, and Shmuel Peleg. Computing occluding and transparent motions. *Int. J. Comput. Vision*, 12(1):5–16, 1994.
- [62] Hailin Jin, Daniel Cremers, Dejun Wang, Emmanuel Prados, Anthony J. Yezzi, and Stefano Soatto. 3-d reconstruction of shaded objects from multiple images under unknown illumination. *International Journal of Computer Vision*, 76(3):245–256, 2008.
- [63] Hailin Jin, Stefano Soatto, and Anthony J. Yezzi. Stereoscopic shading: Integrating multi-frame shape cues in a variational framework. In *2000 Conference on Computer Vision and Pattern Recognition (CVPR 2000), 13-15 June 2000, Hilton Head, SC, USA*, pages 1169–1176, 2000.
- [64] Hailin Jin, Stefano Soatto, and Anthony J. Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *International Journal of Computer Vision*, 63(3):175–189, 2005.

- [65] Hailin Jin, Anthony J. Yezzi, and Stefano Soatto. Mumford-shah on the move: Region-based segmentation on deforming manifolds with application to 3-d reconstruction of shape and appearance from multi-view images. *Journal of Mathematical Imaging and Vision*, 29(2-3):219–234, 2007.
- [66] Neel Joshi and David J. Kriegman. Shape from varying illumination and viewpoint. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–7, 2007.
- [67] Douglas Scott Kay and Donald Greenberg. Transparency for computer synthesized images. In *SIGGRAPH '79: Proceedings of the 6th annual conference on Computer graphics and interactive techniques*, pages 158–164, New York, NY, USA, 1979. ACM.
- [68] Kalin Kolev, Maria Klodt, Thomas Brox, and Daniel Cremers. Continuous global optimization in multiview 3d reconstruction. *International Journal of Computer Vision*, 84(1):80–96, 2009.
- [69] Kiriakos N. Kutulakos and Eron Steger. A theory of refractive and specular 3d shape by light-path triangulation. In *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, pages 1448–1455. IEEE Computer Society, 2005.
- [70] Kiriakos N. Kutulakos and Eron Steger. A theory of refractive and specular 3d shape by light-path triangulation. *International Journal of Computer Vision*, 76(1):13–29, 2008.
- [71] Hendrik P. A. Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph.*, 22(2):234–257, April 2003.
- [72] Jianguo Li, Eric Li, Yurong Chen, Lin Xu, and Yimin Zhang. Bundled depth-map merging for multi-view stereo. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, pages 2769–2776, 2010.

REFERENCES

- [73] Michael Lustig, David Donoho, and John M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
- [74] K. Madsen, H. B. Nielsen, and O. Tingleff. Methods for non-linear least squares problems (2nd ed.), 2004.
- [75] Wojciech Matusik, Hanspeter Pfister, Remo Ziegler, Addy Ngan, and Leonard McMillan. Acquisition and rendering of transparent and refractive objects. In *Rendering Techniques*, pages 267–278, 2002.
- [76] Paul Merrell, Amir Akbarzadeh, Liang Wang, Philippos Mordohai, Jan-Michael Frahm, Ruigang Yang, David Nistér, and Marc Pollefeys. Real-time visibility-based fusion of depth maps. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–8, 2007.
- [77] Ivana Mikic, Pamela C. Cosman, Greg T. Kogut, and Mohan M. Trivedi. Moving shadow and object detection in traffic scenes. In *ICPR*, pages 1321–1324, 2000.
- [78] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):73–82, 2004.
- [79] Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio. Example-based object detection in images by components. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(4):349–361, 2001.
- [80] Nigel J. W. Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. In *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, pages 1573–1580. IEEE Computer Society, 2005.
- [81] Nigel J. W. Morris and Kiriakos N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–8. IEEE, 2007.

- [82] Nigel J. W. Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1518–1531, 2011.
- [83] S. Nadimi and B. Bhanu. Physical models for moving shadow and object detection in video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(8):1079–1087, August 2004.
- [84] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Trans. Graph.*, 24(3):536–543, 2005.
- [85] Ren Ng, Ravi Ramamoorthi, and Pat Hanrahan. All-frequency shadows using non-linear wavelet lighting approximation. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 376–381, New York, NY, USA, 2003. ACM.
- [86] P. Peers and P. Dutré. Wavelet environment matting. In *EGRW '03*, pages 157–166, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.
- [87] Pieter Peers and Philip Dutré. Inferring reflectance functions from wavelet noise. In *Rendering Techniques*, pages 173–182, 2005.
- [88] Pieter Peers, Dhruv Mahajan, Bruce Lamond, Abhijeet Ghosh, Wojciech Matusik, Ravi Ramamoorthi, and Paul E. Debevec. Compressive light transport sensing. *ACM Trans. Graph.*, 28(1), 2009.
- [89] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.
- [90] G. Peyre. Best basis compressed sensing. *IEEE Transactions on Signal Processing*, 58(5):2613–2622, may 2010.
- [91] Andrea Prati, Rita Cucchiara, Ivana Mikic, and Mohan M. Trivedi. Analysis and detection of shadows in video streams: A comparative evaluation. In *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA*, pages 571–576. IEEE Computer Society, 2001.

- [92] Andrea Prati, Ivana Mikic, Mohan M. Trivedi, and Rita Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(7):918–923, 2003.
- [93] Ravi Ramamoorthi, Dhruv Mahajan, and Peter Belhumeur. A first-order analysis of lighting, shading, and shadows. *ACM Trans. Graph.*, 26(1), January 2007.
- [94] Tomás Rodríguez. Shadow removal for robust vehicle detection. *MG&V*, 17(3):249–266, January 2008.
- [95] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [96] Dimitris Samaras, Dimitris N. Metaxas, Pascal Fua, and Yvan G. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *2000 Conference on Computer Vision and Pattern Recognition (CVPR 2000), 13-15 June 2000, Hilton Head, SC, USA*, pages 1480–1487, 2000.
- [97] K. Sato and S. Inokuchi. Three-dimensional surface measurement by space encoding range imaging. *Journal of Robotic Systems*, 2:27–39, 1985.
- [98] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, pages 519–528, 2006.
- [99] Pradeep Sen, Billy Chen, Gaurav Garg, Stephen R. Marschner, Mark Horowitz, Marc Levoy, and Hendrik P. A. Lensch. Dual photography. *ACM Trans. Graph.*, 24(3):745–755, 2005.
- [100] Pradeep Sen and Soheil Darabi. Compressive Dual Photography. *CGF*, 28(2):609 – 618, 2009.
- [101] Pradeep Sen and Soheil Darabi. Compressive estimation for signal integration in rendering. *Computer Graphics Forum*, 29(4), 2010.

- [102] Pradeep Sen and Soheil Darabi. Compressive rendering: A rendering application of compressed sensing. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):487–499, 2011.
- [103] C. H. Séquin and E. K. Smyrl. Parameterized ray-tracing. In *SIGGRAPH '89: Proceedings of the 16th annual conference on Computer graphics and interactive techniques*, pages 307–314, New York, NY, USA, 1989. ACM.
- [104] Sudipta N. Sinha, Philippos Mordohai, and Marc Pollefeys. Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh. In *IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007*, pages 1–8, 2007.
- [105] A.R. Smith and J.F. Blinn. Blue screen matting. In *SIGGRAPH '96*, pages 259–268, New York, NY, USA, 1996. ACM.
- [106] Fredric Solomon and Katshushi Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(4):449–454, April 1996.
- [107] Peng Song, Xiaojun Wu, and Michael Yu Wang. Volumetric stereo and silhouette fusion for image-based modeling. *The Visual Computer*, 26(12):1435–1450, 2010.
- [108] Jiui Sun, Melvyn Smith, Lyndon Smith, Sagar Midha, and Jeff Bamber. Object surface recovery using a multi-light photometric stereo technique for non-lambertian surfaces subject to shadows and specularities. *Image Vision Computing*, 25(7):1050–1057, July 2007.
- [109] Shoji Tominaga and Brian A. Wandell. Color. chapter Standard surface-reflectance model and illuminant estimation, pages 70–78. Jones and Bartlett Publishers, Inc., USA, 1992.
- [110] Philip H. S. Torr and Andrew Zisserman. Feature based methods for structure and motion estimation. In *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, pages 278–294, London, UK, 2000. Springer-Verlag.

- [111] Son Tran and Larry S. Davis. 3d surface reconstruction using graph cuts with surface constraints. In *Computer Vision - ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part II*, pages 219–231, 2006.
- [112] Radim Tylecek and Radim Sara. Refinement of surface mesh for accurate multi-view reconstruction. *International Journal of Virtual Reality*, 9(1):45–54, 2010.
- [113] Joost van de Weijer, Theo Gevers, and Jan-Mark Geusebroek. Edge and corner detection by photometric quasi-invariants. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(4):625–630, 2005.
- [114] Daniel Vlasic, Pieter Peers, Ilya Baran, Paul Debevec, Jovan Popović, Szymon Rusinkiewicz, and Wojciech Matusik. Dynamic shape capture using multi-view photometric stereo. In *ACM SIGGRAPH Asia 2009 papers*, SIGGRAPH Asia '09, pages 174:1–174:11, New York, NY, USA, 2009. ACM.
- [115] George Vogiatzis, Carlos Hernández Esteban, Philip H. S. Torr, and Roberto Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12):2241–2246, 2007.
- [116] Douglas Voorhies and Jim Foran. Reflection vector shading hardware. In *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 163–166, New York, NY, USA, 1994. ACM.
- [117] M.B. Wakin, J.N. Laska, M.F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K.F. Kelly, and R.G. Baraniuk. An architecture for compressive imaging. In *2006 IEEE International Conference on Image Processing*, pages 1273 –1276, oct. 2006.
- [118] Ingo Wald and Philipp Slusallek. State of the art in interactive ray tracing, 2001.
- [119] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the Eurographics Symposium on Rendering Techniques, Grenoble, France, 2007*, pages 195–206. Eurographics Association, 2007.

REFERENCES

- [120] Gregory J. Ward. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 265–272, New York, NY, USA, 1992. ACM.
- [121] Y. Wexler, A.W. Fitzgibbon, and A. Zisserman. Image-based environment matting. In *EGRW '02*, pages 279–290, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.
- [122] Tim Weyrich, Jason Lawrence, Hendrik P. A. Lensch, Szymon Rusinkiewicz, and Todd Zickler. Principles of appearance acquisition and representation. *Found. Trends. Comput. Graph. Vis.*, 4(2):75–191, February 2009.
- [123] WIKI. http://en.wikipedia.org/wiki/Structured-light_3D_scanner.
- [124] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [125] Chenglei Wu, Yebin Liu, Qionghai Dai, and Bennett Wilburn. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *IEEE Trans. Vis. Comput. Graph.*, 17(8):1082–1095, 2011.
- [126] Chenglei Wu, Bennett Wilburn, Yasuyuki Matsushita, and Christian Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 969–976, 2011.
- [127] Chunlin Wu, Jiansong Deng, and Falai Chen. Diffusion equations over arbitrary triangulated surfaces for filtering and texture applications. *IEEE Trans. Vis. Comput. Graph.*, 14(3):666–679, 2008.
- [128] Chunlin Wu, Jiansong Deng, Falai Chen, and Xue-Cheng Tai. Scale-space analysis of discrete filtering over arbitrary triangulated surfaces. *SIAM J. Imaging Sciences*, 2(2):670–709, 2009.
- [129] Chunlin Wu and Xue-Cheng Tai. Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models. *SIAM J. Imaging Sciences*, 3(3):300–339, 2010.

- [130] Chunlin Wu, Juyong Zhang, Yuping Duan, and Xue-Cheng Tai. Augmented lagrangian method for total variation based image restoration and segmentation over triangulated surfaces. *J. Sci. Comput.*, 50(1):145–166, 2012.
- [131] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Computer Vision - ACCV 2010 - 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers, Part III*, pages 703–717, 2010.
- [132] Tai-Pang Wu and Chi-Keung Tang. Photometric stereo via expectation maximization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):546–560, 2010.
- [133] Tai-Pang Wu, Kam-Lun Tang, Chi-Keung Tang, and Tien-Tsin Wong. Dense photometric stereo: A markov random field approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(11):1830–1846, 2006.
- [134] Masaki Yamazaki, Sho Iwata, and Gang Xu. Dense 3d reconstruction of specular and transparent objects using stereo cameras and phase-shift method. In *ACCV (2)*, pages 570–579, 2007.
- [135] Sai Kit Yeung, Tai-Pang Wu, Chi-Keung Tang, Tony F. Chan, and Stanley Osher. Adequate reconstruction of transparent objects on a shoestring budget. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 2513–2520. IEEE, 2011.
- [136] Kuk-Jin Yoon, Emmanuel Prados, and Peter Sturm. Joint estimation of shape and reflectance using multiple images with known illumination conditions. *International Journal of Computer Vision*, 86:192–210, 2010. 10.1007/s11263-009-0222-4.
- [137] Yusuke Yoshiyasu and Nobutoshi Yamazaki. Topology-adaptive multi-view photometric stereo. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 1001–1008, 2011.

REFERENCES

- [138] Tianli Yu, Ning Xu, and Narendra Ahuja. Recovering shape and reflectance model of non-lambertian objects from multiple views. In *CVPR*, pages 226–233, 2004.
- [139] Tianli Yu, Ning Xu, and Narendra Ahuja. Shape and view independent reflectance map from multiple views. *International Journal of Computer Vision*, 73(2):123–138, 2007.
- [140] Andrei Zaharescu, Edmond Boyer, and Radu Horaud. Transformesh : A topology-adaptive mesh-based approach to surface evolution. In *Computer Vision - ACCV 2007, 8th Asian Conference on Computer Vision, Tokyo, Japan, November 18-22, 2007, Proceedings, Part II*, pages 166–175, 2007.
- [141] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape from shading: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(8):690–706, 1999.
- [142] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris N. Metaxas, and Xiang Sean Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, (0):–, 2011.
- [143] Jiayuan Zhu and Yee-Hong Yang. Frequency-based environment matting. In *12th Pacific Conference on Computer Graphics and Applications (PG 2004), 6-8 October 2004, Seoul, Korea*, pages 402–410. IEEE Computer Society, 2004.
- [144] D.E. Zongker, D.M. Werner, B. Curless, and D.H. Salesin. Environment matting and compositing. In *SIGGRAPH '99*, pages 205–214, New York, NY, USA, 1999. ACM Press.

Publication List:

- Qi Duan, Juyong Zhang, Jianmin Zheng, Jianfei Cai: Joint Mesh Refinement and Light Function Recovery for High-quality 3D Reconstruction Under Unknown Illumination. (In Preparation).
- Qi Duan, Jianfei Cai, Jianmin Zheng: Compressive Environment Matting. *Comput. Graph. Forum* (submitted).
- Qi Duan, Jianfei Cai, Jianmin Zheng, Weisi Lin: Fast environment matting extraction using compressive sensing. ICME 2011: 1-6
- Qi Duan, Jianmin Zheng, Jianfei Cai: Flexible and Accurate Transparent-Object Matting and Compositing Using Refractive Vector Field. *Comput. Graph. Forum* 30(6): 1812-1824 (2011)
- Qi Duan, Jianfei Cai, Jianmin Zheng: Vector field fitting for real-time environment matting of transparent objects. ICIP 2009: 3229-3232