

3D Reconstruction using Generalized Perspective Photometric Stereo

Von der Fakultät für MINT - Mathematik, Informatik, Physik,

Elektro- und Informationstechnik

der Brandenburgischen Technischen Universität Cottbus–Senftenberg

zur Erlangung des akademischen Grades eines

Dr. rer. nat.

genehmigte Dissertation

vorgelegt von

M.Sc.

Maryam Khanian

geboren am 26.2.1982 in Kermanshah, Iran

Gutachter: Prof. Dr. rer. nat. habil. Michael Breuß

Gutachter: Prof. Dr. habil. Douglas Cunningham

Gutachter: Prof. Dr. Jean-Denis Durou

Tag der mündlichen Prüfung: 13.7.2018

Abstract

Reconstruction of the 3D shape information is a fundamental problem in computer vision. Among different shape recovering technologies, photometric stereo is highlighted for its capability to produce high quality 3D reconstruction. This dissertation generalizes photometric stereo in different aspects towards creating a practical 3D reconstruction. The proposed techniques can be considered as a fundamental support to develop future cameras offering 3D shapes for various applications such as movie and video game industry, medical sciences, virtual reality, automotive driving and etc. The first generalization is developed for addressing specularities in 3D reconstructions and also involving the perspective projection. These attempts lead to remove the limitation of working with diffuse materials and confined projected scenes. We will prove the applicability of our approach using complex scenes like endoscopy images. In the second proposed approach, we will offer a real-time 3D reconstruction of micro-details with a more generalized reflectance model. Moreover, a recurrent optimization network will be provided. These innovations lead to presenting the 3D reconstruction of details which are even invisible to human eyes like micro-prints on the banknote. This information recovery can be used in various areas such as detecting security items on financial documents for fraud detection and also the quality control of any industrial productions including delicate details such as printed circuits. In the third proposed model, we develop a PS reconstruction technique using neural networks for the uncalibrated PS where the light direction is not available. Finally, for the first time, benefiting from deep neural networks and meta heuristic algorithms, we will devise an approach which can deliver high qualified 3D shape from the internet and out-door images, without any pre-necessary knowledge.

Deutsche Zusammenfassung

Die Rekonstruktion von 3D Gestaltinformationen ist eine fundamentale Aufgabenstellung im Computer-Vision Bereich. Unter den verschiedenen Techniken ist das sogenannte Photometrische Stereo Verfahren (PS) hervorzuheben, da es im Vergleich zu anderen Ansätzen ein höheres Potenzial für eine hoch genaue 3D Rekonstruktion besitzt. Diese Arbeit befasst sich mit Verallgemeinerungen klassischer PS-Ansätze, die dieses Potenzial weiter ausschöpfen sollen. Die Entwicklungen können beispielsweise verwendet werden, um in Zukunft die 3D Information in verschiedensten Bereichen für einen Anwender zu erschließen, wie etwa in der Medizin, der Film- und Videospieleindustrie, beim autonomen Fahren oder für Anwendungen in der virtueller Realität. Die erste der Neuentwicklungen betrifft die systematische Verwendung von hellen Lichtreflexionen in der Berechnung der 3D Rekonstruktion sowie die perspektivische Projektion. Hierdurch werden Beschränkungen üblicher Methoden, wie auf die Rekonstruktion diffus reflektierender Materialien und auf relativ weit entfernte Objekte, aufgehoben. Die Anwendbarkeit des entwickelten Ansatzes wird mittels der Rekonstruktion aus endoskopischen Bildern mit vielen hellen Lichtreflexionen untermauert. In einem weiteren Schritt wird gezeigt, wie man in Echtzeit eine 3D Rekonstruktion auch von sehr feinen Details mittels eines verallgemeinerten Reflexionsmodells erreichen kann. Hierdurch werden hochaufgelöste Rekonstruktionen auch von Details erreicht, die für das menschliche Auge in Bildern unsichtbar sind. Hieraus ergeben sich viele potenzielle Anwendungsmöglichkeiten, etwa im Bereich der automatischen Detektion von Mikrodruck, die im Sicherheitsbereich Verwendung finden, oder in der industriellen Produktion für die Detektion sehr feiner Strukturen wie etwa gedruckter Schaltkreise. Als ein weiterer Beitrag der Arbeit wird ein Verfahren für das sogenannte nicht-kalibrierte PS entwickelt, bei dem neuronale Netzwerke verwendet werden, um die in diesem Fall fehlende Information der Beleuchtungsrichtung auszugleichen. Weiterhin wird zum ersten Mal in der Literatur beschrieben, wie man basierend auf Metaheuristiken und tiefen neuronalen Netzen eine qualitativ hochwertige 3D Rekonstruktion allein mittels Bildern aus dem Internet oder natürlichen Bildern ohne weiteres zusätzliches Wissen erlangt.

Acknowledgments

I am grateful to my supervisor Prof. Dr. Michael Breuß for valuable advice; his generous, kindness, and encouragement impressed me as a Ph.D. student. In addition, I would like to thanks his family (Doris, Sonja, Johannes, Christian, Jonathan, Dominik) for their kindness.

Furthermore, I wish to express my sincere gratitude to Prof. Dr. Douglas Cunningham and Prof. Dr. Jean-Denis Durou for reviewing this thesis.

I am also grateful for my committee members, Prof. Dr. Carsten Hartmann and Dr. Laurent Hoeltgen, for their services and helps.

To my **Father** and **Mother**

My parents are always the source of my spirit and my energy. Without their support in all my life even when I am far from them, I would not be where I am now. Furthermore, I would appreciate all the kindness and the warmth of my husband's parents.

To my husband

Thanks for all your support. You are my motivation for the hard attempt to get our common dreams.

To my brothers

You showed me there is no deadlock for the human.

To Neda and my nephews

I love those moments that I spend with you, those moments that I can forget my tough days. I also thank all the nice people that I have met in Cottbus: Prof. Dr. Ekkehard Köhler and his family for their amiability, Prof. Rembert Reemtsen and his family (Isabelle and David) for all their kindness, Prof. Georg Bader for his considerations and Dr. Friedemann Kemm for his nice statements. My special thanks to our technical colleague Mrs. Silke Büttner that is always ready for any kind of help, and our secretary, Annette Kallweit, as well as our group members, Martin Bähr, Ashkan Mansouri Yarahmadi, Georg Radow and Robert Dachsel.

Journal Publications and Pre-prints

- Maryam Khanian, Ali Sharifi Boroujerdi, Michael Breuß (2018), “Photometric Stereo for Strong Specular Highlights”, *Computational Visual Media*, Springer.
- Maryam Khanian, Ali Sharifi Boroujerdi, Michael Breuß (2018), “Real-Time 3D Reconstruction for Micro Details”, arXiv preprint, arXiv:1802.06140.

Conference Publications

- Ali Sharifi Boroujerdi, Maryam Khanian, Michael Breuß (2017), “Deep Interactive Region Segmentation and Captioning”, *International Conference on Signal Image Technology and Internet-based Systems (SITIS)*.
- Maryam Khanian, Ali Sharifi Boroujerdi, Michael Breuß (2015), “Perspective Photometric Stereo Beyond Lambert”, *International Conference on Quality Control by Artificial Vision (QCAV)*.

Contents

1	Computer vision	27
1.1	Introduction	27
1.1.1	Motivations	31
1.1.2	Challenges	31
1.1.3	Contributions and outline	34
2	From surface to image	37
2.1	Image formation	37
2.1.1	Light source	37
2.1.2	Camera	38
2.1.3	Surface	40
2.1.4	Reflectance model	41
2.1.5	Brightness at a surface point	46
2.1.6	Diffuse reflectance model	48
2.1.7	Dichromatic reflectance model	49
2.1.8	Microfacet theory	50
2.1.9	Computing specular reflection	51
2.1.10	Why Blinn- Phong model instead of Phong model	54
2.1.11	Blinn-Phong reflectance model	55
2.1.12	Cook-Torrance reflectance model	57
2.1.13	Light interaction with the surface	58
2.1.14	Types of light sources	59
2.1.15	Different lighting models	61

2.1.16	The geometric model of perspective projection	63
3	Our approach using the Blinn-Phong reflectance with different perspective projections	69
3.1	Introduction	69
3.1.1	Our contributions	69
3.1.2	Related works	70
3.1.3	Modifying normal vectors	73
3.1.4	Direct perspective surface parameterization	74
3.1.5	Sensitivity of the solution	75
3.2	Perspective Blinn-Phong reflectance model	76
3.2.1	Numerical approach	77
3.3	CCD cameras	80
3.4	Experiments	82
3.4.1	Tests of accuracy	82
3.4.2	Slant and tilt	84
3.4.3	Perspective methods and CCD camera model	88
3.4.4	Tests of applicability on real world test images	89
3.5	Summary and conclusion	98
4	Real-time 3D shape of micro-details	99
4.1	Introduction	99
4.1.1	Our contributions	99
4.1.2	Related works	100
4.1.3	Lightening sensitivity analysis	101
4.1.4	Optimization process	102
4.1.5	Different optimization algorithms	103
4.1.6	Quasi-Newton with BFGS updating	103
4.1.7	Levenberg-Marquardt	104
4.1.8	Powell's Dog Leg	105
4.1.9	Dijkstra Gaussian Mean Curvature (DGMC) technique	105

4.1.10	Recurrent Optimization Network (RON)	106
4.1.11	Integration	107
4.2	Experiments	107
4.2.1	Tests of accuracy on synthetic data	108
4.2.2	Test of applicability on real-world images	112
4.3	Summary and conclusion	113
4.3.1	Algorithms	116
5	A supervised learning approach for uncalibrated PS using a general reflectance model	125
5.1	Introduction	125
5.1.1	Our contributions	125
5.1.2	Related works	125
5.1.3	Artificial Neural Networks	126
5.1.4	The architecture of our proposed method	128
5.1.5	Network specular component	128
5.1.6	Network weight initialization	130
5.1.7	Post-processing step	131
5.2	Experiments	132
5.3	Summary and conclusion	132
6	Deep meta heuristic algorithm for 3D reconstruction	135
6.1	Introduction	135
6.1.1	Our contributions	135
6.1.2	Artificial intelligence	135
6.1.3	Computational intelligence	136
6.1.4	Meta heuristic algorithms	136
6.2	Cuckoo optimization algorithm for 3D reconstruction	139
6.2.1	Brood parasitism for reproduction	139
6.2.2	Cuckoo algorithm	140
6.2.3	Convolutional Neural Networks (CNNs)	142

6.2.4	Deep learning architecture	145
6.2.5	Experiments	146
6.2.6	Summary and conclusion	147
7	Conclusion	157
7.1	Summary	157
7.2	Future directions	158

List of Figures

1-1	Perspective perception of a scene captured by the camera [9].	32
2-1	A simple model of image formation [73]. Light is emitted by the light source and then is reflected from a surface. A portion of this light is directed towards the camera. This process is an imitation from image formation in our eyes.	39
2-2	Different interactions of light with a surface [73].	40
2-3	The geometry of light reflection at a surface [73]. Reflection is generally described by a bidirectional reflectance distribution.	42
2-4	The solid angle is the area of a small patch region on the surface of the sphere [73]. This concept comes about when the plane angle is extended into the 3D space.	44
2-5	We should consider the light flow through a neighborhood of directions to specify the amount of light arriving at or leaving a surface [73].	45
2-6	A small element of surface (a pixel or a surface point) illuminated by a point light source [73].	46
2-7	Incident light arriving from all incoming directions constitute the amount of light reflected towards the camera or viewer [73].	47
2-8	Diffuse reflection which is produced by rough surfaces and tends to reflect light in irregularly in all directions [73].	48
2-9	A statue shows diffuse reflection as slow variations in intensities [73]. . . .	49

2-10	Theory of microfacet. This theory consider specular reflection by assuming that surfaces are made of microfacets as tinny mirrors reflecting incoming light in the mirror direction around the microfacet normal [73].	51
2-11	Specular reflection. The angle of incident light is equal to the angle of reflected light. Each individual ray obeys the laws of reflection and the viewer can see the reflection at only one point [73].	52
2-12	Microfacets which have equal normal (N) and halfway (H) vectors are oriented into viewing direction. These microfacets contribute to the BRDF [73].	54
2-13	Comparision between Blinn-Phong reflectance model and Phong reflectance model. source:[1]	55
2-14	Blinn-Phong model. The incident light ray direction is reflected onto the specular direction around the surface normal [73].	56
2-15	Point light casts light rays in all directions from a single point. The path of light from this source casts a shadow with the uniform density [73].	60
2-16	Extended light source in which the light is not coming from one single point and produces shadow with varying density [73].	61
2-17	Orthographic painting (depicted by a Persian painter) [4]. All objects are seen in the same size without considering their distance and view space (near and far: objects located far away should be projected smaller) and it shows only one side of an object on a principal plane. Furthermore, parallel lines are imaged as parallel lines in contrast to the perspective projection.	64
2-18	Perspective painting of Iranian Imperial Mirror Hall (depicted by the famous Persian artist Kamal-ol-molk) [7] which resembles our eye's view to feel the distance effect (foreshortening), 3D space and shape. Moreover, straight line is imaged as a straight line, whilst others are not. The horizon is the vanishing line of the ground plane.	65

2-19	Perspective viewing of the scene captured by a camera (Zandiyeh complex in Shiraz, Iran) [16]. This projection model is a reasonable close match to the way that an eye or camera lens generates images of objects in our 3D world and make the view more realistic looking.	66
2-20	Perspective projection of the real point R to the image plane $\bar{\Omega}$ [73].	67
3-1	Highly specular photometric stereo setup illustrated by a complex synthetic experiment [73]. In real world, surfaces show both specular and diffuse reflections, so considering only diffuse component or specular component singly is not enough for real world applications. This surface is illuminated by three non-coplanar light sources with both specular and diffuse lights. Shading due to each light is captured in a perspective CCD camera. As can be seen, considering all stated assumptions, we are able to recover shape with high degree of surface details.	72
3-2	Comparison of the surface reconstruction techniques [73]. Top: Input image. Bottom left: Our 3D reconstruction using orthographic projection. Bottom right: Our 3D reconstruction by perspective projection. It can be observed that the perspective approach is able to generate a more compatible result with respect to the original image.	81
3-3	Comparing our results with two described perspective methods regarding their depth reconstructions [73]. Considering all mentioned innovations yields to an accurate perspective PS able to handle specularities.	83
3-4	Set of three test images used for our 3D reconstruction [73]. (a) Real scene used for reprojecting; (b) and (c) are rendered images used for our 3D reconstruction in presence of specularity.	84
3-5	An account of reprojected Beethoven images [73]. Left: Second input image for PS. Middle: Reprojected second image obtained from PPN method. Right: Reprojected second image using the PPS technique.	84

3-6	Two angular variables of slant and tilt [73] as the important parameters used for encoding the 3D perceptions in our brain [109, 125]. The tilt of a surface corresponds to the direction of largest variation in perceived distance and slant varies with the magnitude of the gradient according to (3.38).	85
3-7	Slant (red) and tilt (blue) components of surface orientation on a unit sphere [73]. Red circles (including two bi-colors) represent an area with equal slant angels and blue circles (including two bi-colors) indicate an area with equal tilt angels on the sphere.	86
3-8	Mean square error of tilt angle (MSEA in degrees) for two mentioned perspective techniques applied on the sphere against different focal lengths [73]. As can be seen, the error in estimating tilt angle is decreasing by increasing focal length.	86
3-9	Mean square error of slant angle (MSEA in degrees) for two mentioned perspective techniques applied on the sphere against different focal lengths [73]. The trend of error is descending with respect to gradual increasing of focal length. This amount of decreasing can be seen in PPN perspective projection more than PPS perspective technique.	87
3-10	First and second row: Left: Groud truth. Middle: depth reconstruction from complete Blinn-Phong model with PPN approach. Right: depth reconstruction from complete Blinn-Phong model with PPS approach. These results turn out the proficiency of the proposed method for appealing reconstruction of the images including strong specularities. In addition, PPN approach achieves more faithful reconstructions. Last row: Depth reconstruction from Lambertian model in the presence of specularity accompanied by different perspective projection. Left: PPN approach. Right: PPS method. As it can be seen, the Lambertian model is not able to provide a faithful reconstruction for the specular surface [73].	89

3-11	First row: Four purely specular input images as applied in purely specular model of [89] and the 3D reconstruction of the [89] approach which shows deviations especially around the highly specular areas. Second row: Three ordinary input images including both diffuse and specular components as the input of our method and Our 3D reconstruction. Note that our method does not need the decomposition of the input images into purely diffuse and purely specular components which is a very difficult task even for synthetic images [73].	90
3-12	Test images with high specularity used in realistic real world senario [73]. These images are produced in an endoscopy experiment. So, they do not benefit from any laboratory facilities or confine to the controlled setup conditions.	93
3-13	Depth reconstruction from real world endoscopy images: (first column) results of Lambertian model, (second column) results of the first proposed method (complete Blinn-Phong using PPS) and (third column) results of second proposed approach (complete Blinn-Phong model using PPN). All images are shown from identical view to show the differences. The deviation in the Lambertian results can be clearly seen, while the results of our approach provide faithful 3D reconstruction without any deviation and also with a high amount of details [73].	94
3-14	Depth reconstruction from real world endoscopy images: (first column) results of Lambertian model, (second column) results of the first proposed method (complete Blinn-Phong using PPS) and (third column) results of second proposed approach (complete Blinn-Phong model using PPN). All images are shown from identical view to show the differences. Once more, the deviation in Lambertian outcomes is clear, whereas our approach provides a trustable 3D reconstruction without any deviation [73].	95

3-15	Depth reconstructions from real world images: (a) Results of our proposed method using complete Blinn-Phong model, (b), (c) results of [133]. Both images are shown from identical view to represent the differences. We have also cropped some parts of our results and shown them together with the same cropped area of outcomes of [133] in (c). As it is clear, our approach shows significant superiority over [133] in terms of advantages such as smoothness over the rough output of [133], reconstruction success in specularities and absence of deviation from natural symmetric shape [73]. . . .	96
3-16	Depth reconstructions from real world images: (a) results of our proposed method using complete Blinn-Phong model, (b) results of [133]. As it is mentioned in [133], they could not obtain the reconstruction in the presence of eyes (due to the specularities) unlike our approach which provides faithful results even with including eyes [73].	97
4-1	A successful 3D reconstruction from very fine details of micro-prints on a 50 euro banknote shown in gray color rectangular parts of (A) and (B) [73]. As can be seen, these details are even invisible to human eyes, while our real-time 3D reconstruction from these regions is able to reveal hidden information and features in high amount of details. This information recovery can be used in various applied areas such as detecting security items on financial documents for fraud detection and also the quality control of any industrial productions that include delicate details such as printed circuits. .	101
4-2	A complex real scene including several objects with different sizes captured by a mobile phone with a simple setup which is used for our 3D shape reconstruction [73].	102
4-3	Comparison with [75] and results obtained by Lambertian reflectance (as the most common model applied in PS) in presence of specularity. For a fair evaluation, these experiments are performed against various material parameter (k_s) for different input images [73].	109

4-4	MSE of our 3D reconstructions with different optimization procedures for specular surfaces with varying material parameter (k_s) for different input images [73].	111
4-5	Left: Total MSED and Right: Total MAEN gained by our approach and compared methods on all images mentioned in Table 1 [73]. As it can be seen in the graph, the proposed method can be also favored due to its higher accuracy w.r.t both obtained normal vectors and depth map. To illustrate, our technique can provide accuracy improvement in normal field extraction over [67] by 97.78 % and over [21] by 96.87 % . Furthermore, this improvement in accuracy of depth reconstruction by our scheme over [67] is 89.51 % and over [21] is 76.17 %	112
4-6	Left: input images including specularity [73]. Right: our 3D reconstruction results using perspective Cook-Torrance reflectance model. These results illustrate the capability of the proposed method for providing faithful reconstructions with high frequency details even for the fine details of a tiny surface (e.g. first input image).	119
4-7	First row: input specular images with directional light [73]. Second row: our 3D output using DLPS. Third row: input images with pointwise light. Last row: our 3D output with PLPS. Images produced by directional light are imposed by more specularity and shadows in comparison with images with pointwise light. These shadow parts lead to losing some reconstructed area.	120
4-8	Comparison of several PS methods. From left: 1) input image (Military-man) including high specularities [73], and the 3D results of 2) our reconstruction applying perspective Cook-Torrance reflectance model without any deviation or artifacts, 3) reconstruction of [145], 4) reconstruction of [89] which shows deviations specially around the highly specular areas 5) [21] shows problem in producing complete 3D reconstruction in presence of spcularity and 5) [67] which still includes distortions.	121

4-9	From left: 1) input real-world image including several tiny statues captured without laboratory setting cf. Fig. 4-2, 2) our 3D reconstruction with Cook-Torrance reflectance model, 3) our 3D reconstruction using DLPS for the face of the smallest statue in the scene and 4) our 3D reconstruction using PLPS for the same statue [73].	121
4-10	More real-world experiments with specularity, first row) input images including Einstein statue, two Berlin souvenir statues and a metallic coin, second row) their 3D reconstruction using DLPS and last row) their 3D reconstruction using PLPS. High quality details and structures recovered by the proposed approach confirm offered advantages [73].	122
4-11	Some qualitative evaluations for the proposed method [73]: from left) zooms of our reconstruction results corresponding to the red, blue, green and magenta rectangular areas in Coping as synthetic input and Einstein, Coin and first Berlin statue as the real scenes. It can be seen that our method achieves desirable reconstruction quality even for very fine details.	123
4-12	Reconstructed albedo of two color input images shown in Fig. 10. As can be seen, our RON approach can produce non-uniform albedo from images with a variety of colors. Since the albedo is not same across the entire real-world surfaces, our variant albedo extraction is more adjusted for real-world situations. It should be noticed that these statues are very tiny surfaces [73].	123
5-1	Different components of a biological neuron [10]. The figure is redrawn based on [14]	126
5-2	The general concept of an artificial neuron [10]. The figure is redrawn based on [15]	127
5-3	Architecture of the proposed symmetric neural building block [73].	129

5-4	3D and albedo reconstruction using uncalibrated PS with NN-based learning. First row: input images obtained from [50]. Second row: results of our proposed hybrid neural network model for providing the albedo and depth reconstruction	133
5-5	Comparison with other neural network based 3D reconstructions. First row: input images [50]. Second row: results of the proposed neural network in [35]. Third row: results of the adaptive neural network suggested in [83]. Last row: results of our proposed hybrid neural network model benefiting from a supervised learning and a general reflectance model which shows more improvements regarding the complex albedo and depth reconstruction provided by our model.	134
6-1	Cuckoo mimics host birds eggs [3]. Two top images are examples of a successful mimicking and the down image still needs more attempts to provide more similarities.	141
6-2	A repaint [73] of the general architecture of a Convolutional Neural Network (CNN).	144
6-3	Architecture of the deep model that we used to provide initial values of the depth and normal estimation for the cuckoo algorithm [73].	146
6-4	First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth maps and 3D reconstruction obtained by the proposed deep architecture model. These results gained by one input image without any information about light direction or scene. As can be seen our approach can recover structures successfully which agrees well with the input image.	148
6-5	First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth map and 3D reconstruction obtained by the proposed deep architecture model.	149

6-6	First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth map and 3D reconstruction obtained by the proposed deep architecture model.	150
6-7	First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth map and 3D reconstruction obtained by the proposed deep architecture model.	151
6-8	First row: three input real-world images from MIT Intrinsic Images [55]. Second, third and last rows: our normal, depth map and 3D reconstruction produced by the proposed PS model. These results obtained without any information about the light direction, scene, etc.. As can be seen our proposed metaheuristic deep architecture can provide high qualified results for uncalibrated PS.	152
6-9	First row: three input real-world images from MIT Intrinsic Images [55]. Second, third and last rows: our normal, depth map and 3D reconstruction obtained by the proposed PS model. We think that our approach is successful in producing faithful normals, depth and 3D reconstruction for uncalibrated PS.	153
6-10	First row: three input real-world images from MIT Intrinsic Images [55]. Second, third and last rows: our normal, depth map and 3D reconstruction obtained by the proposed PS model. Our model could reconstruct all these information simultaneously as well.	154
6-11	First row: three input internet real-world images from Moon [2]. Second, third and last rows: our normal, depth map and 3D reconstruction obtained by the proposed PS model (left) and by the deep network (right). It can be seen that applying meta heuristic PS on deep network outputs can improve results specially for tiny details. These results obtained without any information about the light direction, scene, etc.. Astronomical images can be regarded as a challenging experiment, our technique is able to provide reasonable results even for this experiment.	155

List of Tables

3.1	Comparison between MSE of the reprojected Beethoven images from two described perspective methods of PPN and PPS.	83
3.2	MSE of the reconstructed depth from images with specularities by two perspective methods of PPN and PPS. As it is clear, we consider 3D reconstruction in the presence of both diffuse and specular reflection simultaneously from the surface which leads to involving both k_d and k_s and applying complete Blinn-Phong model. In addition, we applied both diffuse and specular light. Finally, we extended our model to different perspective projection techniques.	88
3.3	MSE of the reconstructed depth from images with high specularities shown in Figure 3-11.	90
4.1	MSE of our 3D shape reconstruction with perspective Cook-Torrance reflectance model in the presence of specularities using different lightening models.	110
4.2	Our total speed improvement over other methods for all images of Table 4.3. It can be seen that our scheme can decrease consuming time dramatically as well as providing reliable results.	110
4.3	Comparison of several PS methods w.r.t Mean Square Error of Depth (MSED), Mean Angular Error of Normal vectors (MAEN in degrees), and Time in seconds. Innovations aggregated in our approach leads to a significant improvement regarding obtained normal vectors and depth as well as computational Time.	115

Chapter 1

Computer vision

1.1 Introduction

In computer vision as the science of automatic analysis of useful information, our aim is doing an inverse operation. It means that we try to describe the world that is seen in one or more images and to reconstruct its characteristics such as shape, translucency, illumination, color and texture distributions. It is fascinating that humans and animals do this task very easily and unconsciously using their neural resources, while computer vision techniques are so prone to error because it is an ill-posed problem due to the loss of depth information and inferences. People who have not experienced this field may underestimate difficulties in this area. For example, providing a software which is able to recognize and name all the people in an image is very challenging and hard task. This misinterpretation that vision should be easy dates back to the early days of emerging artificial intelligence, when it was initially believed that the cognitive (logic proving and planning) parts of intelligence were intrinsically more difficult than the perceptual components [26]. The good point is that computer vision is very popular in a wide variety of real-world applications. Some of them can be mentioned as:

- **Biological information:** Studying and processing of our biological vision system using some visual mechanisms from retina to cortex for different tasks such as mimicking the behaviour of biological systems at different levels of complexity, biometric-

based visual identification of people and inventing various robotic helpers. This also led to some of the learning-based methods e.g. image and feature analysis or classification based on the neural net and deep learning approaches.

- **Quality control:** Providing automatic inspections to ensure the quality of products or looking for defects. To illustrate, by exploring a series of physicochemical characteristics of goods such as its maturity, size, weight, shape, color, the existence of dirt, diseases, stem, seeds, sugar content, etc. These features can cover all components influencing the appearance of a product and properties related to its conservation.
- **Object recognition:** This task can be applied to different fields such as automated checkout lanes, security monitoring, alerting and detection of the anomaly.
- **Medical imaging:** This topic includes a wide range of processing techniques in medical sciences. For example, registering pre-operative and intra-operative imagery, performing long-term studies of brain morphology, intelligently aiding the surgeon, producing data to provide us with the better diagnosis, treatment and prediction of diseases.
- **3D modeling :** Automated construction of 3D models from images used in different systems such as Bing maps, movie and video game industry, medical sciences to provide detailed models of organs, architectural models, designing of new devices and vehicles, the earth science community to construct 3D geological models and physical devices that are built with 3D printers or Computer Numerical Control (CNC) machines.
- **Automotive safety:** As the study and design of equipment for detecting unexpected obstacles to minimize traffic collisions, accidents, improving road safety, developing driver assistance and self-driving cars.

This extensive range of applications is due to this fact that computer vision systems present a considerable amount of information about different attributes of scene analyses. In addition, this technology provides the possibility of studying scenes in the level of the electromagnetic spectrum in which the human visual system is not sensitive, such as ultraviolet

radiation (UV) or infrared (IR) spectral regions.

Our rich human visual system is capable to perceive diverse essential information about the three-dimensional (3D) configuration of objects, using different cues. This 3D impression as the shape of real-world objects is done automatically by our brain when we are looking at the environment. Much of our 3D vision can be even perceived already from 2D images. A common question for researchers is whether a computer vision system can provide this brain ability for a variety of applications such as modern future cameras, scene understanding and describing, industrial modeling, robotics, virtual reality, etc. Inspired by the evolution of capturing devices and many modern trends one may conjecture that the need of reliable and fast methods for 3D shape reconstruction is going to increase and this may represent an important building block of future mainstream multimedia applications. Understanding human cognition is also another increasing interest in analyzing 3D shapes of facial surfaces with various applications such as Human-Computer Interaction (HCI), facial surgery, biometrics, 3D animation and video communications.

A classification of 3D shape from X techniques can be provided based on the structure of these techniques. They can be divided into two types of approaches: geometric based methods and photometric based techniques. Geometric approaches such as multiview stereo and structure-from-motion give sparse 3D points based on finding correspondence points on images and triangulation operation on projection rays.

Many input images taken from different viewpoints and constant illumination should be provided for these approaches (please refer to [113] for a survey). The detailed quality of produced 3D points depends on the performance of correspondence matching because the similarities in appearance of multi-view images make difficulties for matching corresponding points [38]. For further reading on stereo vision (cameras calibration, stereo matching algorithms, reconstruction, etc.), we refer the reader to a research provided in [128]. In general, multiview stereo and structure-from-motion focus on obtaining the rough depth (coarse shape information) using geometric constraints instead of the highly detailed surface. On the other hand, photometric approaches like Photometric Stereo (PS) and shape from-shading offer a dense normal field using shading information. This shading information are caused by shading variations in images due to the interactions of shape (surface

normal orientation), reflectance and lighting.

To illustrate, we can consider a rectangle white plane and a cylindrical bottle which has a rectangle shape in its front side with an equal size of the rectangle white plane. Taking images from both objects with the same viewpoint produces two rectangles of the same sizes. But we can recognize the differences and the 3D volume of the bottle since the image of plane presents constant reflected light intensity; whereas the bottle shows brighter intensity in areas closing to the light source (and vice versa). This basic idea of shading information as the gray level changes are considered as changes in the direction of the surface normal which display the 3D shape.

Shape from focus is another scheme for 3D shape reconstruction. This technique uses a sequence of images taken by a camera at different focus levels to compute the shape of objects. The performance of shape from focus depends on an effective estimating of the focus. Specular or translucent materials can make problems for the focus estimation and produce inaccurate results.

Shape from shading [63] can also provide 3D information. It uses one input image for producing a 3D shape. Unfortunately, it is an ill-posed problem. Because different objects which are illuminated from different lighting directions can produce the same image appearance. So, the normal of a patch cannot be obtained uniquely from its intensity and there are infinite solutions in the case of using a single image. Applying more images as utilized in PS, can solve this problem as the consecutive observed intensities for the surface patches lead to only a single normal orientation.

In PS several input images taken from a fixed view point under different illumination directions are used for the 3D reconstruction. The pioneers of the problem are Woodham [144] [64] and Horn et al. [65] . As shown by Woodham [145] the orientation of a Lambertian surface can be uniquely determined provided at least three input images are given. Since in PS the pixel-wise surface normal map in the same resolution of the input image is produced, it reaches a high accuracy which is not accessible by any geometric approach. So PS is a potential technique when a graceful geometric structure of objects is required and it provides the estimation of surface properties rather than image properties.

PS has received wide attention in different practical applications such as industrial quality

control [45, 154, 97, 70], face recognition [151, 155, 153, 50], medical science [100, 138], texture classification [111, 121, 88] and estimation of weather condition [117]. Another benefit of PS is to generate invariant features utilized in different kinds of recognition tasks [132, 122].

1.1.1 Motivations

Most of the later works on PS kept some simplifying building blocks. Therefore, obtaining reliable PS results is still a challenging task when dealing with real-world images. For example, they consider the Lambertian surface reflectance for their 3D reconstruction, following the Lambert's law [79]. However, it may easily fail for surfaces whose reflection is non-Lambertian. Corresponding effects e.g. highlights as occurring on shiny objects are quite readily observable in real-world images. The orthographic projection is another considered assumption for the camera model. But the latter assumption is realistic when objects are far away from the camera, but not if they are close since then perspective effects grow to be important. Using a single point light source distantly (distant or directional lighting) is also a general assumption which is not commonly occurred. In addition, the main limitation is that the light source direction should be given in advance; whereas for the complex case no information is available about the lighting setup. In addition, another trend is the widespread use of smart phones. Consequently, it can be expected that multimedia applications based on real-time 3D technology with inexpensive devices can constitute an important research area. Considering all aforementioned points, the key problem is how to generalize the assumptions of photometric stereo and we focus on this topic to significantly extend scenarios for the wide applicability of PS as well as providing real-time 3D reconstruction.

1.1.2 Challenges

We consider here our framework for the generalized PS as the PS setup does not strictly meet the mentioned simplified assumptions in the previous section. We can conclude this section, summarizing all the hypotheses as:



Figure 1-1: Perspective perception of a scene captured by the camera [9].

Light source

Different types of light sources can be defined based on the emission from a single point in space with some varying distribution of outgoing lights. At first, we consider the distant light, also known as a directional light as an emitter which deposits illumination from the same direction at every point in space. As well as directional lightening, they do not attenuate their energy with the distance between the light source and object [101]. Starting from the distant light source, we will equip our models with the spatially varying lightening and consider point-wise light source.

Camera model based on the perspective projection

Projection simply means creating a view of an object on a plane and in the real-world, our eyes perceive scenes as the perspective projection (cf. Fig. 1-1). So, it is important to implement the camera model in 3D reconstruction approaches as the perspective camera where depicted objects look more like what our visual system realizes and not as the orthographic one.

Surface reflectance

Geometry is a part of the image formation process. In order to generate an image, we should determine not only how to project a surface, but also what brightness value should be assigned to it. So, the interaction of light with the surface is at the core studies in computer vision. In many of the proposed works for 3D reconstruction, the reconstructed object is supposed to be Lambertian surface which leads to considering the reflectance function based on the Lambert's law. The latter case is observed on diffuse objects like a white wall. However, objects in real-world strictly admit the Lambertian assumption for reflectance, and even many diffuse materials such as matte plastic, wood, fabric, etc. have deviations from this simplest assumption. Therefore, we will address PS with more realistic reflectance functions.

Neural Network (NN) based PS for uncalibrated light direction

The main point of PS is that the light source direction should be given in advance as a pre-knowledge as many other methods also need some necessary given information. On the other hand, this information is not always available. Consequently, we design a neural network structure not to only solve the problem of uncalibrated PS and also to leverage the hierarchical properties of neural networks.

Internet and out-door images

A successful 3D reconstruction should be able to provide 3D reconstruction from ordinarily available images such as the internet or out-door images where there is no any information about the scene, light direction or depth values and image is the only available information. So, we will devise a deep meta heuristic algorithm for dealing with these difficulties which is able to provide outstanding results in these complex scenes.

Real-time 3D reconstruction

By emerging 3D scanners, a new age for industrial activities and also commercial companies is provided. Despite their popularity, they can obtain only the rough geometry of scanned objects due to the limited assumption that the scanned surfaces should be Lambertian. So, they break down in the presence of specularities for non-Lambertian objects.

Therefore, offering a real-time accurate 3D scanner to recover fine details of a scanned object is still a challenge. Following the mentioned necessity, and also motivated by the growing demand for interactive environments, we equipped our approach with a real-time per-pixel process in addition to generalizing photometric stereo in different aspects where it can be significantly applicable.

Providing consumer-level equipment

Image capture process should be possible with inexpensive camera equipment and instruments. Specialized equipment like professional situation provided in laboratory should not be required. Instruments should be easy available so that it would be sufficient for general consumers. Our real-world images used for our 3D reconstruction are provided without any laboratory facilities.

Practical input data capture

The image capture process should be easy, straightforward and not time-consuming. As few images as possible should be required. Calibration objects should be avoided. It is clear that increasing the number of input images leads to the complexity of the image formation process and is time-consuming.

To summarise: the ideal technique uses inexpensive equipment, requires few input images and minimal or no calibration. In addition, the resulting 3D reconstruction per pixel is fast to compute and provides high level accuracy. These points motivated us to use the minimum necessary input images in all our experiments captured by an ordinary camera like smartphone camera in a simple room and not in a laboratory. Finally, in the last chapter, we will extend our approach to work with even one internet input image.

1.1.3 Contributions and outline

The main contributions of this dissertation are summarized as follows: **Methods for photometric stereo with general reflectances:**

Since the Lambertian assumption is not always accurate in real cases and pure Lamber-

tian objects can seldom be found in real-world, more general reflectances are considered to extend the generality of the materials that can be handled. Consequently, we explore non-lambertian reflectance models for obtaining an accurate 3D shape from surfaces with non-Lambertian effects.

Real-time photometric stereo with perspective projection using recurrent optimization network:

We facilitate models with the perspective projection which is a manner closest to how we perceive the objects with our eyes. Furthermore, a real-time 3D reconstruction is provided. In addition, we propose our recurrent optimization network to obtain an accurate 3D reconstruction of micro details.

Adopting more light setting in PS :

In addition, we adjust our approaches for a different light set up which can be applied in different situations.

Neural network based PS:

Furthermore, we design a practical photometric stereo system by benefiting from the neural network framework that works without knowing light directions and using a general reflectance model. This model would be of great value in obtaining 3D surface representation.

Deep uncalibrated PS using a meta heuristic technique:

In addition, we will provide a deep PS using a meta heuristic technique. This approach can provide highly qualified 3D shape from out-door or internet images without any pre-necessary information like light direction, scene information or depth values.

The attempts and achievements in this dissertation alleviate the classical assumptions and promote the practical competences of photometric stereo technique.

This thesis consists of six chapters. Chapter 1 is introductory, it includes a review of existing related 3D reconstruction techniques and assumptions on the properties of the acquisition system and surface which they serve as the basis of their approach. Chapter 2 describes the foundations of the image formation process and its relation to the photometric stereo algorithms. In Chapter 3 we explain the proposed methods which are suitable for the generalization of the photometric stereo to the reflectance properties of much more abundant

types of objects. Chapter 4 presents a real-time 3D reconstruction of micro-details. The proposed algorithm of neural network based photometric stereo is presented in Chapter 5. Thanks to this algorithm, it is possible to obtain the 3D shape without prior knowledge on lighting. Chapter 6 provides meta-heuristic optimization algorithm based on deep learning for uncalibrated PS and internet images.

Chapter 2

From surface to image

2.1 Image formation

The shape of a surface, its reflectance, and the incoming illumination make the image of the object in our eyes or on a camera sensor. The shading changes based on the local orientation of the surface and also the characteristics of incoming light. The ability of the human visual system is remarkable in deducing different information from images while recovering such information is a much more difficult task for a computer system. Since the photometric stereo is a process that drives the information on the surface from the intensity, we will describe this process in this chapter. The image formation process involves the interaction of several essential elements in the scene. We analyze here the relation between image irradiance, light source, camera and surface. Let us discuss in details each factor of this interaction and the acquisition system for images.

2.1.1 Light source

Light source is the first important element of the image formation model since there is not any image without light. When light interacts with surfaces on a macroscopic level, it can be absorbed, transmitted, or reflected.

Infinite light (directional light)

Illumination used during each image acquisition can be constituted from different light sources. The light source used for each image acquisition can be of a directional nature. This kind of light source is of a very small size and can be considered like a point light source at infinity (e.g. distant source like the sun which can be treated as an area light source or the reflected light from the moon). The illumination incident on the object is uniform in direction and intensity. The intensity of the directed light source is related to the number of photons which is emitted by the light source. Color of the source also depends on the wavelength of the emitted light. The emitted light can be in a visible range of wavelengths (white light), or the wavelengths of ultraviolet or infrared ranges.

Spatially point light

Another situation that we will consider in this thesis is developing a technique that is able to consider lighting that varies spatially across the object. This light source casts rays in a cone shape from a single point source. To implement spatially point light, we should consider a geometric model of the object and estimate the indirect illumination which results from the spatially varying incident illumination. Then, we can calculate each incident ray of light in a scene.

In our approach we make an assumption of a single light source, which may vary in direction and intensity from one acquisition to another. This light source is of a visible wavelength range. In the following sections where we want to involve different lightening models in our approach, we will discuss more about light sources.

2.1.2 Camera

Another element of the image formation is the camera. The camera is designed to capture the reflected or emitted light. It can work with the light in the visible spectrum or with other portions of the electromagnetic spectrum. The camera as an optical device produces the image of objects or scenes and records them on an electronic sensor or photographic

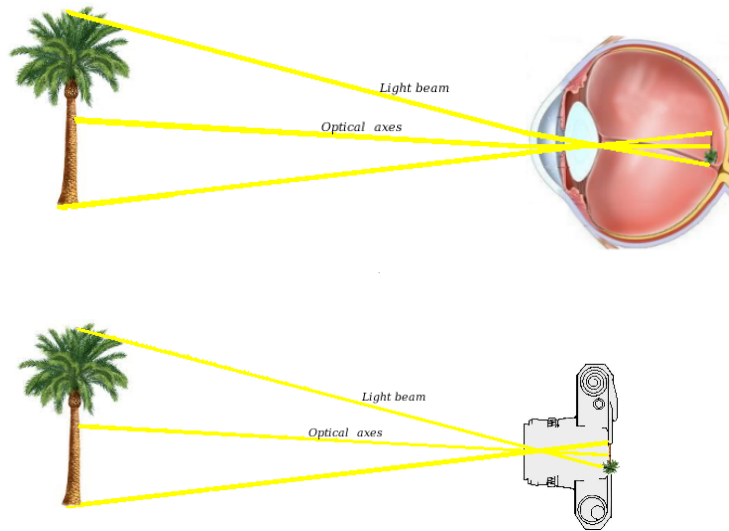


Figure 2-1: A simple model of image formation [73]. Light is emitted by the light source and then is reflected from a surface. A portion of this light is directed towards the camera. This process is an imitation from image formation in our eyes.

film (mostly for the old cameras). The main mechanism for producing an image is imitated from our eyes (cf. Fig. 2-1). This process is defined as follows: light enters into a box through a converging lens and the image is recorded on a medium which is sensitive to light (mainly from a transition metal halides). A shutter is used which controls the length of time for entering the light to the camera. Many of cameras allow the photographer to view the scene and modify it such that the desired area is in focus, or the exposure can be managed to prevent the cases of too bright or too dim. A display screen (usually from a Liquid Crystal Display (LCD)), provides the scene view before recording and settings different parameter likes ISO speed, exposure, and shutter speed. Since we develop the perspective projection in our approach we do not have to put the photographed object far from the camera like in the hypothesis of orthographic projection. The camera is considered to be fixed in the image acquisition leads to the total correspondence between pixels of different images. So, there is no need to deal with the additional problem of point correspondence. In addition, we will model a Charge-Coupled Device camera (CCD camera) in our photometric stereo technique.

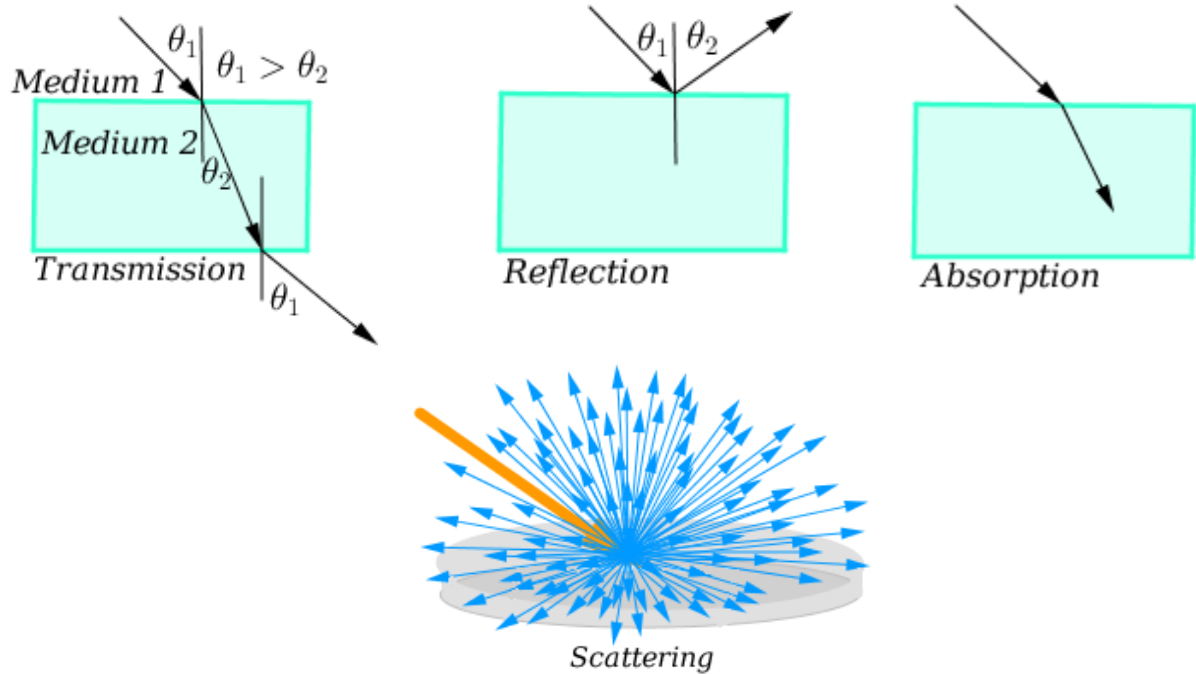


Figure 2-2: Different interactions of light with a surface [73].

2.1.3 Surface

The surface is the most complex component of the image formation process. Once the light interacts with the surface point, the complicated light-matter process occurs. This interaction process depends on the physical characteristics of the light and the physical composition and characteristics of the surface. A part of the incident light is reflected, some of the light is transmitted, and another portion of the light can be absorbed by the medium itself. Since light is an energy form, so the law of energy conservation for light implies that: the light incident at a surface is a summation of light reflected and light absorbed and light transmitted. Reflection is an external surface process. Refraction (or transmission) is an internal surface event and is related to the transparent or translucent object. It is modeled by the means of the Bidirectional Transmittance Distribution Functions (BTDFs) and describes how much light is transmitted when interacting with a material. The BTDF is not used alone for the image formation modeling, it usually applied in combination with Bidirectional Reflectance Distribution Functions (BRDFs) for the transparent objects.

Scattering is composed of both external and internal phenomenon and is the complex in-

cident light interaction with the surface. In this situation, it is supposed that the incident light is transported inside of the material of the surface and interacts with the material and goes out from the material but not necessarily at the point of incidence when it hits the surface. So, the scattering is a combination of transmission and reflection. It is formulated by the Bidirectional Surface Scattering Reflection Distribution Functions (BSSRDFs) as the compound of the BRDFs and the BTDFs. These phenomena are schematically presented in Fig. 2-2. We consider here only the fraction of light that is reflected.

when light hits an object which is a Lambertian surface (an ideal matte), it will be reflected in lots of different directions. This kind of reflectance model has a simple reflectance function and is proportional only to the cosine of the incident angle. In addition, these Lambertian surfaces look equally bright from any direction when they are illuminated under uniform or collimated illumination. Because the amount of light reflected from a unit area goes down as the cosine of the viewing angle, but the amount of area seen in any solid angle goes up at the reciprocal of the cosine of the viewing angle. Consequently, the realized intensity of a surface element is constant with respect to the viewer direction. On the other hand, most surfaces have a specular component in their reflection and in general, some light is reflected at all angles in decreasing amounts of the specular angle. In order to achieve this effect, the cosine of the angle between the predicted specular angle and the viewing angle should be also involved. More discussion about these reflectance models will be provided in following chapters.

2.1.4 Reflectance model

In this section, we present a concise description of a basic radiometric image formation model. Different models have been proposed to model the interaction of light and surface. We first describe the most general form and then deal with some more specialized models, we also discuss how these models can be extended to the different perspective projection techniques and also to the different illumination corresponding to a scene.

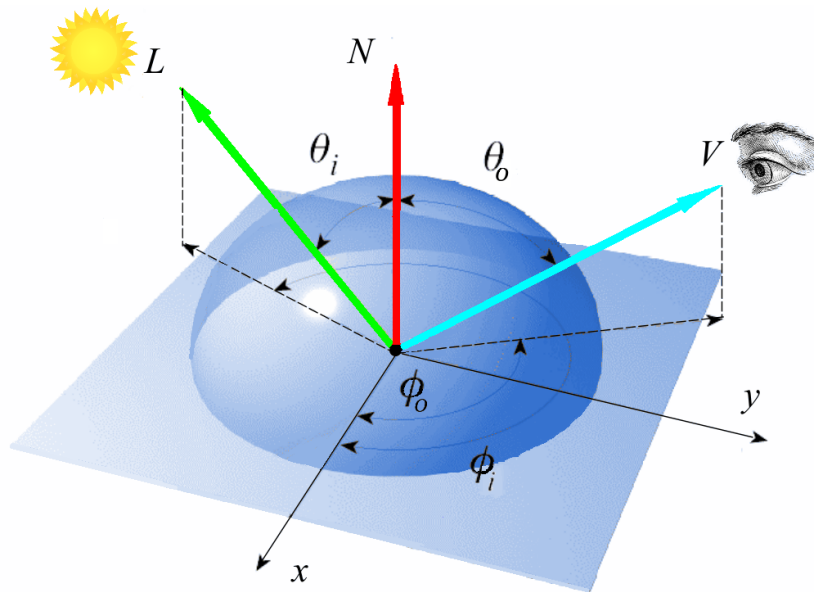


Figure 2-3: The geometry of light reflection at a surface [73]. Reflection is generally described by a bidirectional reflectance distribution.

Bidirectional Reflectance Distribution Function (BRDF)

The BRDF describes surface reflection at a point. Considering the setting in Fig. 2-3, we would like to know how much radiance is outgoing the surface in the direction toward the viewer, as a result of incident radiance.

In general, the extent to which light is reflected depends on the position of viewer and light relative to the surface normal. To illustrate, a shiny surface is illuminated by a point light source. Since it is made of the shiny material, some surface regions show a shiny highlight seen by an observer. If the observer changes his position (changes view direction), the position of the highlight will also change. Similarly, if the light source is moved, the highlight shifts.

Since the BRDF describes how light is reflected, it should consider this dependency on the view and light directions. So, the BRDF is a function of incoming light direction and outgoing view direction relative to a local orientation at the light interaction point. Additionally, when light interacts with a surface, different wavelengths (colors) of light can be reflected in varying degrees (upon the physical properties of the material itself). Therefore, the BRDF is also a function of wavelength λ .

Furthermore, the BRDF can be position variant and different from one surface point (x, y)

to another because light interacts differently with different regions of a surface. This property, known as positional variance and this manner of light leads to the generating of surface details. For example, the ringing and striping patterns which can be seen in wood imply that the BRDF changes with the surface spatial orientation. We consider these dependencies for BRDF and also simplify the representation of the BRDF as $BRDF_{\lambda}(\theta_i, \phi_i, \theta_o, \phi_o, x, y) = BRDF(\theta_i, \phi_i, \theta_o, \phi_o)$, where (θ_i, ϕ_i) and (θ_o, ϕ_o) are respectively the incident and the reflected light directions in the spherical coordinate system for the unit vectors. In the following sections, the BRDF keeps its dependency to the incident light wavelengths, even if its notation is not shown or mentioned explicitly in the given descriptions and explanations. The BRDF is defined as the ratio of radiance L_o to the irradiance E_i which are the quantity of light reflected in the considered direction (θ_o, ϕ_o) and the quantity of light arriving the surface with the incoming direction (θ_i, ϕ_i) :

$$BRDF(\theta_i, \phi_i, \theta_o, \phi_o) = \frac{L_o}{E_i} \quad (2.1)$$

Differential solid angle

As BRDF measures how light reflects off a surface when viewed under different viewing positions, we should have a good perception about the amount of light arrives at a surface element (or leaves a surface element) from a particular direction. So, it is important to describe the concept of a differential solid angle.

In order to deal with the light arriving (or leaving) a surface, it is better to consider the quantity of light arriving at or going across an area of space. The reason for this is that light is measured in terms of flow through an area because the light is an energy per-unit surface area (Watts/meter²). Therefore, we should be careful about applying the amount of light arriving from a direction and it is more meaningful to regard the light coming from a small region of directions.

We can consider a solid angle like 2D angle in a plane (planar angle) which is extended to an angle on a sphere (space angle). So, the solid angle locates in a 3D unit sphere instead

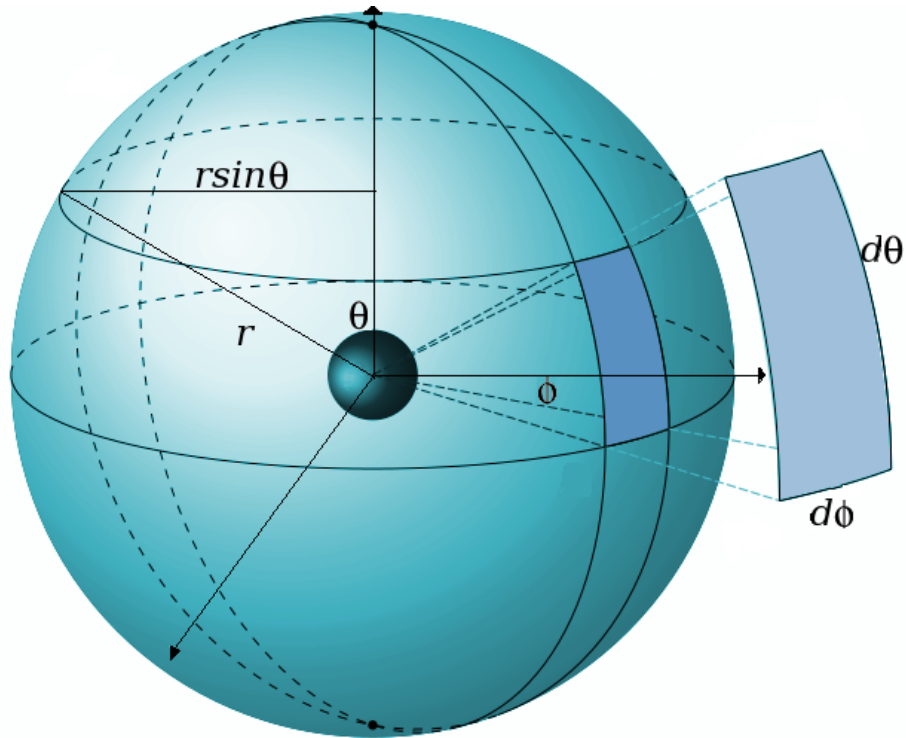


Figure 2-4: The solid angle is the area of a small patch region on the surface of the sphere [73]. This concept comes about when the plane angle is extended into the 3D space.

of a 2D unit circle. The total area s corresponding to the area of a small rectangular region on a unit sphere is the solid angle shown in Fig. 2-4. This is a simple way to understand the concept of solid angle.

In Fig. 2-5 a light direction arrives at a small surface element and a small cone of surrounding incoming directions are presented. The amount of light received by a small surface element can be determined through a small cross-sectional area surrounding a direction. This cone constituting a volume of directions can make the pyramid patch shape of the solid angle on the inside of the sphere like in Fig. 2-4. Since the differential solid angle is regarded as the area of this small patch, we will compute this area in spherical coordinates (θ, ϕ) using small differential angular changes $(d\theta, d\phi)$. The differential solid angle dw is defined as

$$dw = (\text{height})(\text{width}) \quad (2.2)$$

$$dw = (d\theta)(\sin\theta d\phi) \quad (2.3)$$

$$dw = \sin\theta d\theta d\phi \quad (2.4)$$

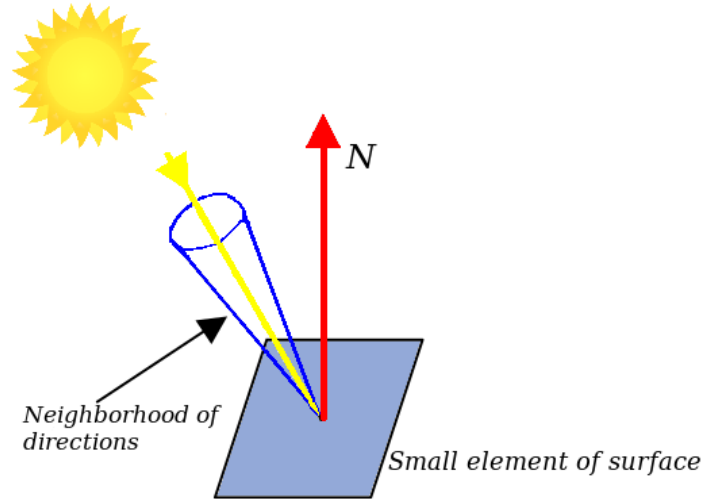


Figure 2-5: We should consider the light flow through a neighborhood of directions to specify the amount of light arriving at or leaving a surface [73].

The width and the height of the rectangular patch are measured in radians. So, the area quantity has the unit of radians squared (or steradians with the abbreviation of sr). In practice, the simplest imagination concept of solid angle is to think of it as the area of a small surface region uniquely defined along with each direction like the small area on the unit sphere defined by a neighbourhood around a given direction.

Now we consider Fig. 2-6. The figure represents a small surface element like a pixel or a point on the surface which is illuminated by a point light source. The amount of light arriving from direction w_i is proportional to the amount of light arriving at the differential solid angle. We show the light source intensity with l_i . Since the differential solid angle as a flat region on the hemisphere is small, it is uniformly illuminated and receives the same quantity of light, l_i , for each position on the differential solid angle. As a result, the total amount of incoming light into the region is $l_i * w_i$.

If we want to consider the amount of light spread on (projected onto) the element of the surface instead of the solid angle, this projection should be done by accounting $\cos\theta_i = N \cdot w_i$ (where θ_i is the angle between the unit vector in direction w_i and the normal vector N) and considering the foreshortened area of the surface element as $\cos\theta_i dw_i$. So, the total energy reaches the surface element dw as the irradiance is:

$$E_i = l_i \cos\theta_i dw_i. \quad (2.5)$$

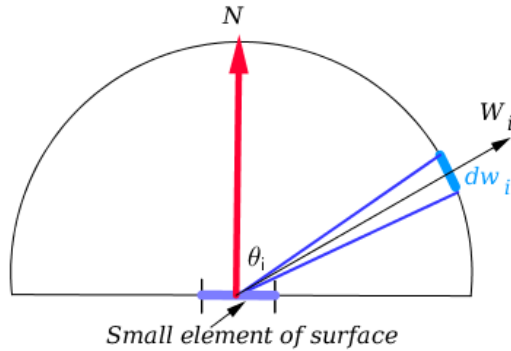


Figure 2-6: A small element of surface (a pixel or a surface point) illuminated by a point light source [73].

So, the quantity of light receiving by the surface element is the intensity of the light times the width of the cross sectional surface area (differential solid angle) on the unit sphere where the light passes.

$$BRDF = \frac{L_0}{l_i \cos \theta_i dw_i}. \quad (2.6)$$

Based on the formulation of BRDF, it is deduced that the BRDF is not bounded to the interval of [0,1]. This is because of the division by the cosine term in the denominator which may leads to values larger than 1 for BRDF. In addition, in the definition of BRDF there is a division by the solid angle with the unit of steradians (sr). So, the unit of a BRDF is the inverse of steradians (sr^{-1}).

2.1.5 Brightness at a surface point

It can be seen that the BRDF is the starting point for the analysis of the surface reflectance properties. Although it can be used to reconstruct surface normal field or to render realistic images, but dealing with it in the presented form is not easy. It is a function dependent on four variables of the angles of the incident and reflected directions $(\theta_i, \phi_i, \theta_o, \phi_o)$. So, it is better to get an easier reflectance function which encodes the main behavior of the surface and obtain the illumination produced at a point of the surface.

In fact, all lights that hit an object point from the hemisphere of incoming directions are involved in the produced illumination (cf. Fig. 2-7) . So, in order to calculate the amount of light generated at a surface point, we should sum the irradiance over all hemisphere using

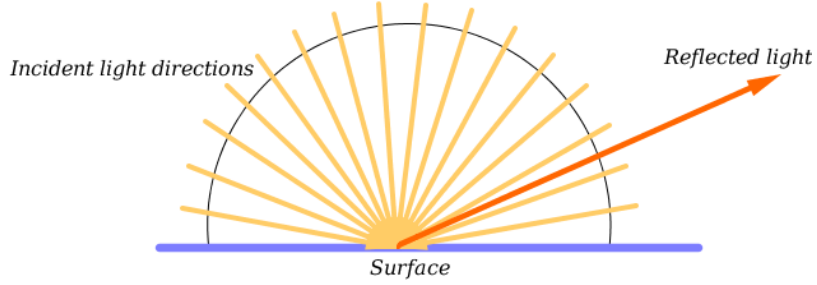


Figure 2-7: Incident light arriving from all incoming directions constitute the amount of light reflected towards the camera or viewer [73].

integration:

$$L_o = \int_{\Omega} L_{oi}(w_i, w_o) dw_i. \quad (2.7)$$

where $L_{oi}(w_i, w_o)$ is the amount of light reflected along with the direction w_o resulting from incident direction w_i . Ω is the hemisphere of incident light directions. If we consider a discrete space rather than in a continuous one (the light sources are discrete like a finite number of point light sources), we can replace the integral with a summation:

$$L_o = \sum_i L_{oi}(w_i, w_o) \quad (2.8)$$

On the other hand, by considering equation (2.6), we know that the amount of reflected light as the radiance resulting from the light arriving with direction w_i , is defined based on the BRDF and irradiance E_i as follows:

$$L_{oi} = BRDF \times E_i \quad (2.9)$$

Since dw_i in the formulation of E_i can be neglected in the discrete case (all incoming directions are equally weighted), so we will have $E_i = l_i \cos \theta_i$. This implies that:

$$L_{oi} = BRDF \times l_i \cos \theta_i \quad (2.10)$$

Equation (2.10) provides the light reflected in the direction of an observer for a single point light source, which is also known as the BRDF brightness equation or BRDF lighting equation [64].

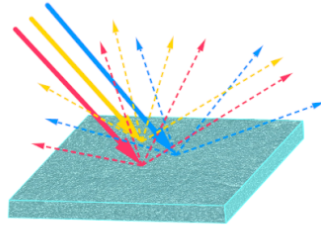


Figure 2-8: Diffuse reflection which is produced by rough surfaces and tends to reflect light in irregularly in all directions [73].

2.1.6 Diffuse reflectance model

The diffuse reflection (known also as the Lambertian or matte reflection) considers the interaction between the surface and light as the way light scatters as the microscopic variations. It means that incident light reflected uniformly in all directions as shown in Fig. 2-8 and shows the smooth variation of intensity like the case that can be seen in the statue Fig. 2-9. Diffuse reflection conveys the body color to the light because it is created by the absorption and redistributing of light inside the material of the object [114, 51].

For this kind of reflectance, the BRDF is a constant value because light is reflected uniformly in all directions and a Lambertian surface looks equally bright from every viewing direction. Therefore, in this case, we will have:

$$BRDF(\theta_i, \phi_i, \theta_o, \phi_o) = BRDF \quad (2.11)$$

it is also shown by ρ and called Albedo (surface chromaticity or body colour). In this dissertation, we denote albedo by the k_d as the diffuse material reflectance. So, equation (2.10) as the light reflected from the surface is obtained as:

$$L_{oi} = k_d \times l_i \cos\theta_i \quad (2.12)$$

where θ_i is the angle between the surface normal N and the incoming light L . By substituting the value of $\cos\theta_i$ in equation (2.12) and removing i from the light source intensity



Figure 2-9: A statue shows diffuse reflection as slow variations in intensities [73].

l_i (we have just one point light source), the following equation will be presented:

$$I = k_d \times l \left(\frac{N \cdot L}{\|N\| \|L\|} \right) \quad (2.13)$$

Where I is the image intensity (the amount of light reflected towards a viewer or camera named the outgoing radiance in the view direction). This equation is also known as the image irradiance equation or reflectance function.

2.1.7 Dichromatic reflectance model

Motivation

When reflection from a surface is specular in nature, highlights are observed. They correspond to large spikes in image intensity value. The Lambertian model is unable to accurately represent specular reflection. As a result, surfaces and textures which exhibit specular reflectance (which are ubiquitous in the environment) are neglected in the Lambertian reflectance, while specularities should be explicitly considered in the reflectance function. It is quite well proved that a light source illuminating most surfaces, reflects a significant part of the light as described by a non-Lambertian reflectance model [129, 28, 23].

Since the assumption considered by Lambertian reflectance is not always accurate for real surfaces, it is not suited for surfaces with complex reflectance behavior. This makes problems not only for a reasonable 3D reconstruction but also a reliable retrieval of the features becomes very difficult or even impossible [66]. As a result, more general reflectance should be considered to extend the generality of the materials that can be dealt with. Two major light reflection components including specular reflection as the high frequency variations and diffuse reflection as the low frequency component should be combined to provide a successful reflectance model [103, 93, 143] since, in reality, most of the objects show both of these reflections in different areas. Therefore, they include both reflection models at the same time.

Consequently, we address the dichromatic reflectance models. The model is named dichromatic because an interface reflection component (specular) and a body scattering component (diffuse) are used to present the reflected light and components described with a geometrical and a spectral term [114]. In addition, it is shown that using a second such term in a BRDF model can lead to decreasing the fitting error by more than 20% [94, 42]. Another attractive point about the dichromatic reflectance model is that they are able to interpret a pixel color constitution regarding the spectral distribution power and the geometrical scale factor [114].

2.1.8 Microfacet theory

This theory assumes that surfaces which are not perfectly smooth are composed of many very tiny facets. This surface is essentially a height field (cf. Fig. 2-10), where the distribution of facet is described statistically. So, the main component of these models is an expression for the distribution of microfacet normals enabling them to effectively model many real surfaces [101, 61]. Each facet is regarded to be a perfect reflector and satisfies the physical laws of reflection for dielectrics. In addition, interreflection of light from several facets before leaving the surface and occlusion of that facets, due to their orientations, are other points considered in this theory [78]. The following reflectance models, addressed in our approach, admit this theory.

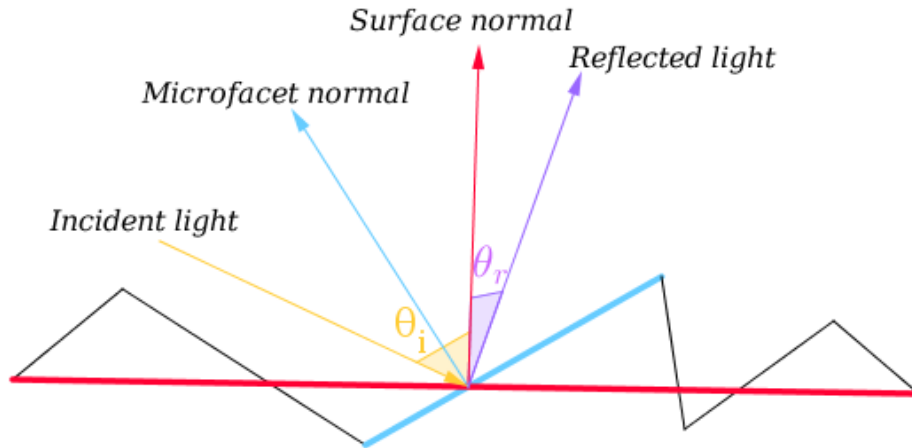


Figure 2-10: Theory of microfacet. This theory consider specular reflection by assuming that surfaces are made of microfacets as tinny mirrors reflecting incoming light in the mirror direction around the microfacet normal [73].

2.1.9 Computing specular reflection

As the second major component of a BRDF is the specular reflection, some insight into the physical modelling of specular reflection will be provided. It will be showed that the specular reflection has a strong dependency on the direction of the outgoing light.

Snell's law

Specular reflection applies Snell's Law as:

1. The incoming ray, the surface normal, and the reflected ray all lie in a common plane.
2. The angle that the incoming ray forms with the surface normal as the angle of incidence θ_i is determined by the angle that the outgoing ray forms with the surface normal as the reflected angle of θ_r .

Fig. 2-11 shows the geometry of specular (perfect mirror) reflection which is derived from Snell's law (second law: law of reflection). For an ideal reflector, such as a mirror, based on the Snell's law, the angle of incidence is equal to the angle of specular reflection, as shown in Fig. 2-11. R is the direction of specular reflection (it is also called a perfect reflecting direction) and V is the viewer direction (which reflected around the normal vector). In the case of the mirror reflection, the specular reflection is visible only when V and R coincide. However, in the case of real objects (not perfect reflectors) the specular reflectance can be

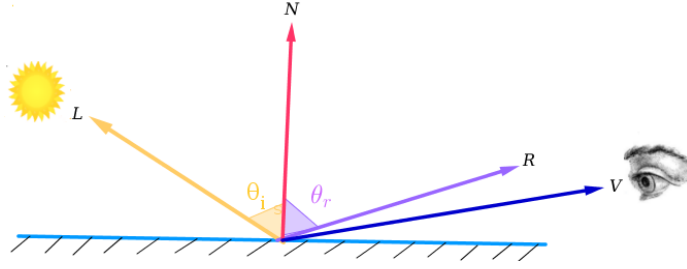


Figure 2-11: Specular reflection. The angle of incident light is equal to the angle of reflected light. Each individual ray obeys the laws of reflection and the viewer can see the reflection at only one point [73].

seen even if V and R do not coincide means that it is visible over a range of values (a cone of values), because the specular reflection may spread out slightly in a cone and light is most intense in the center of cone and falls off in its neighborhood. So, if the viewer is not looking exactly at R , he still observes a reduced reflection.

We compute the specular direction R based on the following assumption:

- $\theta_i = \text{incidence angle}$.
- R is the vector of perfect reflection and is coplanar with L and N .
- $\|N\| = \|L\| = \|R\| = 1$.
- $L.N = R.N = \cos \theta_i = \cos \theta_r$.

We can get the following linear combination for R based on the mentioned condition of coplanar for L , N and R :

$$R = \alpha L + \beta N \quad (2.14)$$

Then taking dot product of N for the equation (2.14) leads to the following expressions:

$$R.N = \alpha L.N + \beta N.N \implies \cos \theta_i = \alpha \cos \theta_i + \beta \quad (2.15)$$

By squaring (2.14), we can get:

$$R.R = \alpha^2 + \beta^2 + 2 \alpha \beta (L.N) \quad (2.16)$$

\implies

$$1 = \alpha^2 + \beta^2 + 2 \alpha \beta \cos \theta \quad (2.17)$$

And finally solving (2.15) and (2.17) results in:

$$\alpha = -1, \beta = 2 \cos \theta = 2L.N \quad (2.18)$$

Substituting α and β in equation (2.14) provides the specular reflection R as:

$$R = 2 (L.N)N - L \quad (2.19)$$

By estimating the specular reflection, the Phong model [102] is presented as:

$$I(x, y) = \underbrace{k_d \left(\frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|} \right)}_{diffuse \ term} l_d + \underbrace{k_s \left(\frac{R(x, y) \cdot V(x, y)}{\|R(x, y)\| \|V(x, y)\|} \right)^n}_{specular \ term} l_s \quad (2.20)$$

k_d is the diffuse material parameter (diffuse color). k_s is the specular material parameter and shows the fraction of specularly reflected light. This is so that specularities appear to show the color of the source as the highlight color because the specular reflection coefficient k_s is not a function of wavelength. $I(x, y)$ is the intensity at pixel (x, y) . We show the diffuse intensity as l_d . In addition, we considered l_s which is the specular light source intensity. Finally, the exponent n is also called the specular exponent or glossiness which determines the sharpness of specularity peaks and varies the rate of falloff. In the Phong model, the strength of specularity is proportional to the angle between V and R .

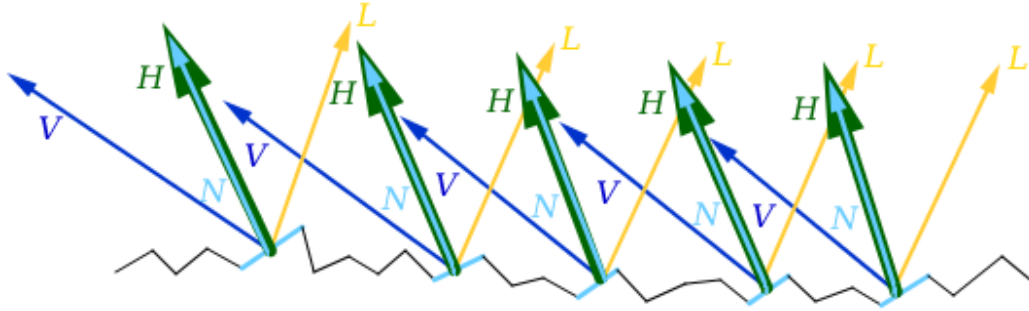


Figure 2-12: Microfacets which have equal normal (N) and halfway (H) vectors are oriented into viewing direction. These microfacets contribute to the BRDF [73].

2.1.10 Why Blinn- Phong model instead of Phong model

The Phong model allows to represent not only the diffuse surface material but also glossy materials with specular component, but we now concentrate on the specular term modeled by Phong reflectance equation. Jim Blinn proposed a modification on the Phong model using Blinn-Phong reflectance model [25].

He used the more physically meaningful vector of H which is defined to be the halfway vector between the viewing vector V and the light vector L . In addition, the half vector is the direction to which the microfacet normals need be oriented to reflect L into V [61] (see Fig. 2-12). While the reflection vector used in Phong model [102] has no such physical significance [61]. He used this approximation that when V overlaps R , H overlaps with N , so the center of reflected energy is still at R .

The justification for this is that H specifies the direction that the surface normal should be so that V would be the mirror direction and the maximum highlight happens and H is the direction of maximum highlight [5]. So, the Blinn-Phong Model uses the halfway vector H , which is faster to compute than reflection vector R and still view dependent since H depends on V .

Now the question is that:

Does using $N.H$ vs. $R.V$ affect highlights? Although Blinn-Phong model is more physically meaningful than the original Phong model, we can ask whether this makes any practical difference for production shading. Yes, highlights spread as seen in Fig. 2-13.

Blinn-Phong may be considered more realistic as the specularly behaves differently, when

the viewer looks at a grazing angle towards the surface. The specularity will be more oval shaped, so it is projected circularly, instead of being circular and projected as an oval. This is more realistic than Phong model [61] and it creates more precise models of empirically determined BRDFs than Phong for many types of surfaces [94]. Ngan et al. [94] explored some underlying reasons for this and showed that BRDFs based on the halfway vector yield more visually plausible results than BRDFs based on the incident direction [42]. Another advantage of the halfway vector is that, in a microfacet-based BRDF model, the halfway vector H equals the normal vector of the microfacets responsible for reflection [42] as seen in Fig. 2-12.

2.1.11 Blinn-Phong reflectance model

The Blinn-Phong reflectance model can be considered as the bi-polynomial model by setting coefficients that are related to the dot product of $N.H$ as a description for the distribution of the specular zone intensities. Therefore, the Blinn-Phong model is a dichromatic model which provides the possibility of representing not only diffuse surface material but also shiny materials with the specular component.

In order to modelling the major component of specular reflection that depends strongly on the direction of outgoing light, Blinn-Phong reflectance model is presented as:

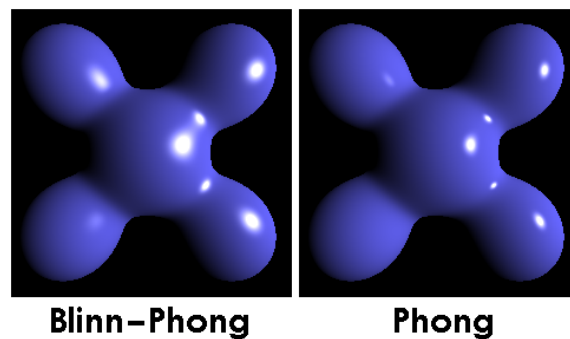


Figure 2-13: Comparison between Blinn-Phong reflectance model and Phong reflectance model. source:[1]

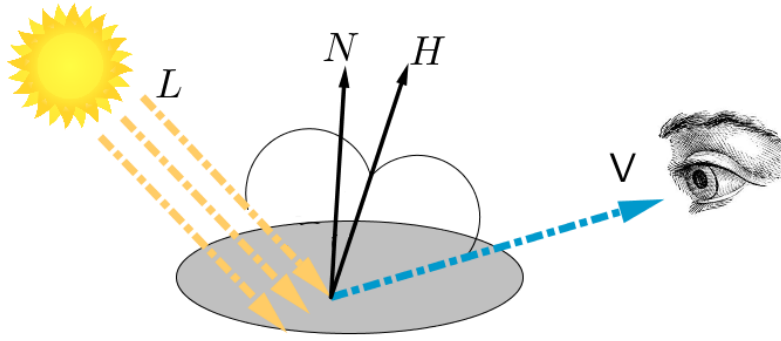


Figure 2-14: Blinn-Phong model. The incident light ray direction is reflected onto the specular direction around the surface normal [73].

$$\begin{aligned}
 I(x, y) = & \underbrace{k_d \left(\frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|} \right)}_{\text{diffuse term}} l_d \\
 & + \underbrace{k_s \left(\frac{H(x, y, z) \cdot N(x, y)}{\|H(x, y, z)\| \|N(x, y)\|} \right)^n}_{\text{specular term}} l_s
 \end{aligned} \tag{2.21}$$

The vector H will coincide with the surface normal if the reflected light coincides with the viewer direction.

The statistical distribution of facet normals in Blinn-Phong model is approximated by $(H \cdot N)^n$ as an exponential falloff. This term can be considered as a lobe projected on the half-vector [33]. One of the most important points regarding the dichromatic reflectance model is the dimensions added to BRDF, means the spatial ones. BRDFs naturally vary with position on a surface, and it is important to model this variation. While this popular variation is provided by Blinn-Phong model, as mentioned before it is not possible to provide it for the case of Lambertian reflection. All elements included in Blinn-Phong model is illustrated in Fig. 2-14.

2.1.12 Cook-Torrance reflectance model

Although Blinn-Phong reflectance model is much more physically meaningful than Lambertian model or the original Phong model but we want to try some more microfacet theory and deal with a model derived from geometrical optics which consider more features like subsurface scattering or the Fresnel reflection in the reflectance model.

Cook-Torrance model [37] like the recent popular theoretical models consider the assumption that the large scale BRDF is the result of fine scale roughness of the surface [110] and the roughness (as a collection of microfacets with random sizes and orientations) is involved in this model. This property enables them to effectively model many real-world surfaces [101] in which, mirror-like facets is modelled by the Cook-Torrance reflectance to retrieve the local shape from specularity. In addition, this model assumes that the facets might be of size comparable to the wavelength of light. This means that the wave nature of light, and hence phenomena such as interference, must be taken into account.

The final model for the BRDF is complex, but agrees well with actual measured BRDFs [110]. Following, the Cook-Torrance model is presented as:

$$\begin{aligned}
 I(x, y) = & \underbrace{k_d \frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|}}_{diffuse\ term} l_d \\
 & + \underbrace{\frac{G(x, y, z) D(x, y, z) F(x, y, z)}{4 \frac{V(x, y, z) \cdot N(x, y)}{\|V(x, y, z)\| \|N(x, y)\|} \frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|}}}_{specular\ term}
 \end{aligned} \tag{2.22}$$

where

$$G(x, y, z) = \min \left\{ 1, T(x, y, z), R(x, y, z) \right\} \tag{2.23}$$

and

$$T(x, y, z) = \frac{\frac{2H(x, y, z) \cdot N(x, y)}{\|H(x, y, z)\| \|N(x, y)\|} \frac{V(x, y, z) \cdot N(x, y)}{\|V(x, y, z)\| \|N(x, y)\|}}{Q(x, y, z)} \tag{2.24}$$

$$R(x, y, z) = \frac{\frac{2H(x,y,z) \cdot N(x,y)}{\|H(x,y,z)\| \|N(x,y)\|} \frac{L \cdot N(x,y)}{\|L\| \|N(x,y)\|}}{Q(x, y, z)} \quad (2.25)$$

$$Q(x, y, z) = \frac{V(x, y, z) \cdot H(x, y, z)}{\|V(x, y, z)\| \|H(x, y, z)\|} \quad (2.26)$$

$$D(x, y, z) = \frac{1}{\pi m^2 \cos^4(\alpha)} \exp\left(\frac{-\tan^2(\alpha)}{m^2}\right) \quad (2.27)$$

$$\alpha = \arccos\left(\frac{H(x, y, z) \cdot N(x, y)}{\|H(x, y, z)\| \|N(x, y)\|}\right) \quad (2.28)$$

$$F(x, y, z) = f_\lambda + (1 - f_\lambda) \left(1 - \frac{H(x, y, z) \cdot V(x, y, z)}{\|H(x, y, z)\| \|V(x, y, z)\|}\right)^5 \quad (2.29)$$

m is the root mean square slope of the surface microfacets (i.e. the surface roughness). It is worth to mention that in expression (2.29) H is used instead of N as suggested in [61]. F is the Fresnel term (to describe the phenomenon that specular is stronger at grazing angle) and f_λ is called the reflection coefficient or refractive index (the reflectance of the surface at normal incidence). Here, the distribution function of facets is described by D and measures the proportionate number of facets oriented at an angle α forming the average normal to the surface and finally G is the geometrical attenuation term. The specular component is assumed to come from the reflection in those facets oriented in the direction of half vector H as also shown in Fig. 2-12.

2.1.13 Light interaction with the surface

Lighting is one of the main component interacting with other elements in image formation. In this section, we present briefly some intuition behind light interactions with the surface. Then we formulate the light models applied in our approach. The question about the iden-

tity of light has been debated for many centuries. The sun is the most important source of radiating light. Electrical instruments are also developed to brighten our darkness, and many other applications of light affect our lives daily. Light is a special kind of electromagnetic energy. It travels quite fast with the speed which is in a vacuum expressed as $c = 2.99 \times 10^8 \text{ m/s}$. Light movement in a vacuum is at a constant speed which is considered a universal constant. It should be mentioned that its speed changes with travelling through non-vacuum media such as air by 0.03 percent slower or glass by 30.0 percent slower. In order to work with light in real applications, we represent light in terms of its magnitude and direction. As it is discussed, reflection as one of the phenomena of light interacting with a surface is explicitly modeled by BRDF which considers all properties of the light source, the object surface and the viewer and quantifies how much light is reflected in the camera direction when it interacts with a surface made of a certain material. The field of detection and measurement of light energy is called radiometry.

2.1.14 Types of light sources

Light sources can generally be divided into point and area light sources.

Point light sources

We begin with the analyzing of point light sources. In computer graphics, we usually treat lights as rays emanating from a source. A point light source originates at a single location in space. Point light source is a simplified model of any small light sources compared to the scene or distant light source such as a distant light bulb, a star, or perhaps the sun, in which we treat it as an infinitesimal. Although, light source is not infinitely small, but assuming that it is a point leads to the simpler analyses. Let us mention that spot lights used for example in theater and film are also supported as a special form of point light which casts rays in a cone shape from a single point source and emits illumination in all directions.

Light sources are related to an output power. It is the amount of energy consumed per unit of time (for example, joules per second). The most common unit of power for light

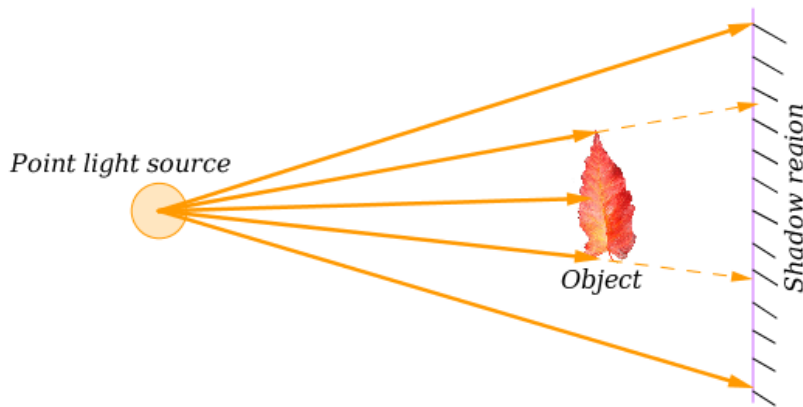


Figure 2-15: Point light casts light rays in all directions from a single point. The path of light from this source casts a shadow with the uniform density [73].

sources is watt such as 60-watt or 100-watt light bulbs. This power valuation is the amount of power consumed by the bulb. In general, it is assumed that the input power (as the consumed energy) and the output power (as light) are the same.

To illustrate, the wattage of the sun is around 3.846×10^{26} watts. A point light source of W watts radiates the same amount of energy in all directions. If a point light source is encompassed by a sphere with the radius of 1 meter, then all of the light output from the source land on the sphere. The emitting power of point light source, or its radiance, is usually measured in watts per unit solid angle or watts per steradian. It may be also simply given in watts, which can be thought of as watts per 4π steradians (the solid angle subtended by all space on the surface of a unit sphere equals 4π).

In practice, instead of using units of radiance for describing the light arriving at a point due to a light source, radiant intensity is used as the proper unit [101].

Extended light sources

Another modelled light sources are extended light sources which are the light source to some extent, whether it is small or large. As a way of example, the entire sky on an overcast day, the light coming through a window on a cloudy day or a simple area light source such as a fluorescent ceiling light. In this type of light, the power of the light emitting from a single infinitesimal point is considered to be negligible or 0. On the other hand, the power emanating from a finite area is of interest. So, in order to calculate the power for extended

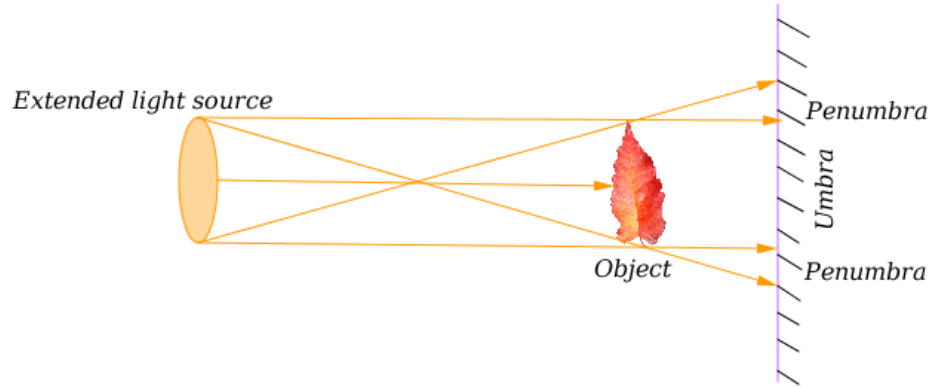


Figure 2-16: Extended light source in which the light is not coming from one single point and produces shadow with varying density [73].

light sources, we consider not just as watts per steradian, but rather, as watts per steradian per unit area. The area refers to the area of the emitter, not to the surface on which light is hitting. Fig. 2-15 and Fig. 2-16 are depicted to simulate light emitting from mentioned light sources.

2.1.15 Different lighting models

The diverse types of light sources can be modeled for illuminating the scene.

Initially, we consider the distant light and then we develop our perspective models with spatially varying lightening by considering point-wise light source.

Directionally varying illumination or isotropic light source

When the surface is lit by a point light source at infinity (like sun), the implication is that the resulting rays of light are parallel with each other when they hit the surface and rays emanate in all directions so that it is named directional light. So, the illumination incident on the scene is considered to be uniform in direction and intensity and we can consider it as an isotropic point light source (the light source have a directional falloff or dependence). Since their energy does not depend on the distance between the light source and object [101], we can represent the light with only a direction vector shown with $L = (\alpha, \beta, \gamma)$.

Spatially varying lighting

So far, many techniques in 3D reconstruction, consider the lighting environment as a directional lighting. While such a measurement registers the directionally varying illumination for light sources and surfaces, it does not take into account the spatially varying illumination in the scene which implies that how the light varies from one point to another. This makes a necessity to model the spatially as well as directionally varying illumination for the scene which simulate the effects of lighting varies spatially across the subject and allows for the independent control of intensities [139]. As a result, building on the base of light, a more complex light based on spatially varying radiance field will be introduced and we show that how such more detailed measurements of lighting can be used to achieve a higher level of realism in 3D reconstructions.

This modelling is partly based on the ray-tracing algorithm. Ray tracing algorithms on computers track the path of light source rays through the scene until they intersect a surface. This approach provides a method for finding the visible surface points seen from any position and direction [101, 118]. That is, knowing which point in space projects onto which point on the image plane allows one to directly associate the radiance at the point to the irradiance of its image [85].

In the case of ray tracing, following the reflected rays and finding their intersection points with the surface is part of the process. Similarly, for spatially varying lighting, we trace the direction of vectors which are the light vectors and try to model their intersections with the surface. It is worth to mention that we also developed the perspective projection in geometric modelling of spatially varying lighting.

To this aim, a three dimensional vector fields for the light vector of L should be imposed instead of a directional light vector L . Furthermore, the rate at which light falls off with distance is another issue which should be induced to make the model more consistent to the actual physical aspects.

Integrating all the mentioned points, the new spatially dependent light vector L is defined as:

$$L(x, y) = \Pi \frac{\Upsilon - S(x, y)}{\|\Upsilon - S(x, y)\|} \quad (2.30)$$

with the light attenuation Π :

$$\Pi = \frac{l_d}{\|\Upsilon - S(x, y)\|^2} \quad (2.31)$$

where Υ is the light source position.

It should be remembered that since we developed the perspective projection, so the perspective surface parameterization is imposed in $S(x, y)$ for this point light source. As a result, the light direction depends not only to the point (x, y) , but also to the depth of the surface itself. So, in this case, an initial guess for the depth of the surface is required.

2.1.16 The geometric model of perspective projection

This part provides some tutorial for introducing the central ideas in the perspective geometry of 2D space and 3D space and how this geometry may be represented or estimated. Since the Renaissance period, paintings with extremely accurate perspective were developed. It is like an accompaniment that early attempts to establish the rules of perspective originate from skilled artists in architecture and engineering. In contrast, orthographic comes from the Greek word for straight writing or drawing. The orthographic transformation captures a rectangular area of the scene and projects it onto the front face of the box defining the area. It does not consider the effect of foreshortening for objects which leads to becoming smaller on the image plane when they get farther away and feeling like it has less depth. In addition, it keeps parallel lines as parallel (collinearity) and it preserves relative distance between objects (cf. Fig. 2-17).

As can be seen in Fig. 2-17 orthographic is single side view (front, left side, right side, top or bottom) of an object, but the perspective is a 3D view as shown in Fig. 2-18, where we can see 3 sides for example, top, left, front of an object which provides for the viewer an understanding of its shape. This reconstruction conducted by such a painting is done where the human's eye view or camera as depicted in Fig. 2-19 is the most common inspiration for sketching.

The perspective projection also projects a volume of space onto a 2D image plane by con-



Figure 2-17: Orthographic painting (depicted by a Persian painter) [4]. All objects are seen in the same size without considering their distance and view space (near and far: objects located far away should be projected smaller) and it shows only one side of an object on a principal plane. Furthermore, parallel lines are imaged as parallel lines in contrast to the perspective projection.

sidering the effect of foreshortening.

An excellent presentation of the algebraic and matrix representations of perspective cameras can be found in [57]. Now, we need to specify how 3D primitives are projected onto the image plane. A mathematical model for this process should account for the following types of transformations:

- *coordinate transformations between the camera frame and the world frame*
- *projection of 3D coordinates onto 2D image coordinates*

These concepts are shown in Fig. 2-20. In order to project the real-world point R to the

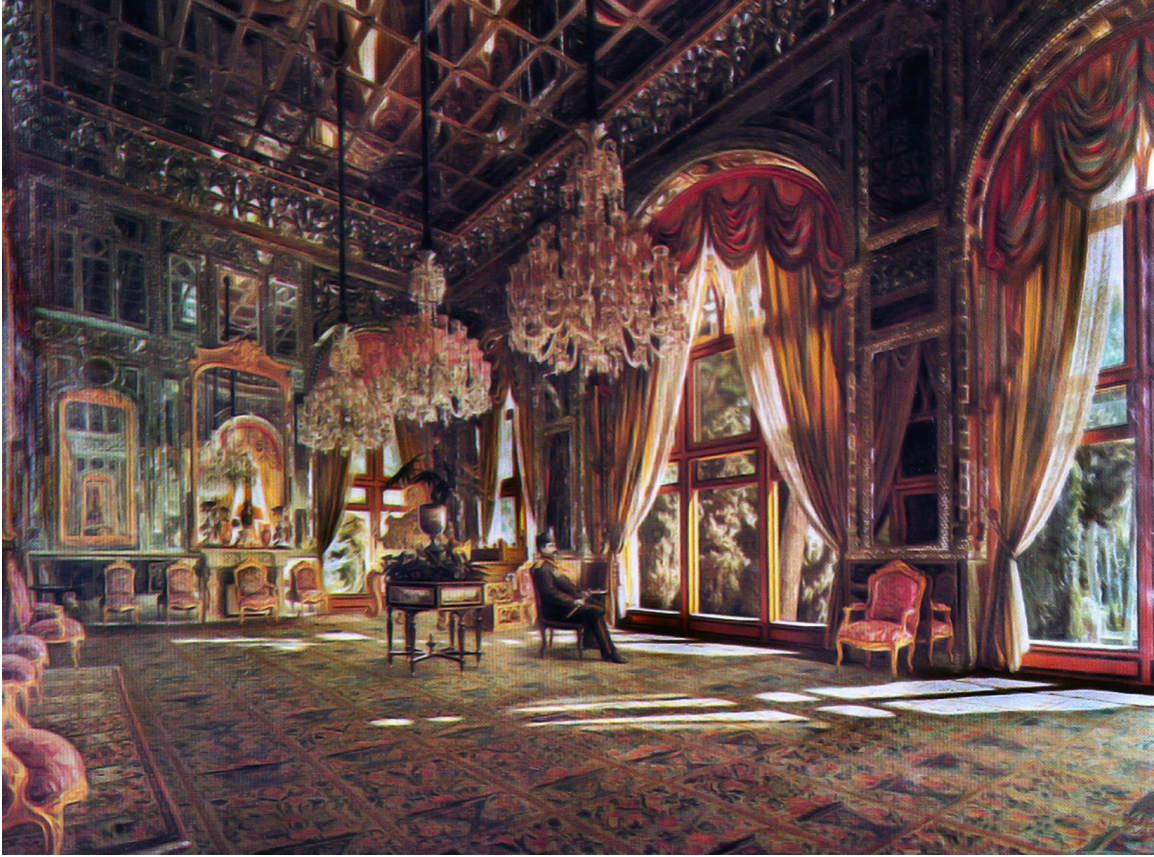


Figure 2-18: Perspective painting of Iranian Imperial Mirror Hall (depicted by the famous Persian artist Kamal-ol-molk) [7] which resembles our eye's view to feel the distance effect (foreshortening), 3D space and shape. Moreover, straight line is imaged as a straight line, whilst others are not. The horizon is the vanishing line of the ground plane.

point r on the image plane Ω , we will consider the Thales theorem in both horizontal red and vertical blue triangles:

$$\frac{f}{z(x, y)} = \frac{x}{u} = \frac{y}{v} \quad (2.32)$$

On the other hand, in reality the image plane Ω lies behind the lens. So, under the perspective projection model, the surface parameterization is performed as follows:

$$S(x, y) = \begin{bmatrix} \frac{-xz(x, y)}{f} \\ \frac{-yz(x, y)}{f} \\ z(x, y) \end{bmatrix} \quad (2.33)$$

where f is the focal length.



Figure 2-19: Perspective viewing of the scene captured by a camera (Zandiyeh complex in Shiraz, Iran) [16]. This projection model is a reasonable close match to the way that an eye or camera lens generates images of objects in our 3D world and make the view more realistic looking.

The normal vector as the cross product of the partial derivatives of the surface is computed as:

$$N(x, y) = \begin{bmatrix} \frac{z(x,y)}{f} \nabla z(x, y) \\ \frac{z(x,y)}{f^2} (\nabla z(x, y) \cdot (x, y) + z(x, y)) \end{bmatrix} \quad (2.34)$$

where $\nabla z(x, y) = (z_x, z_y)$ is the gradient field. This more accurately models the behavior of real cameras.

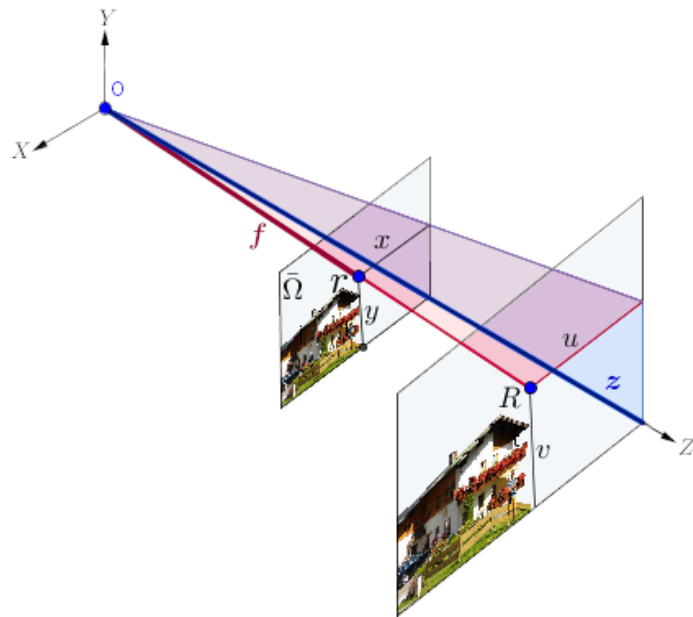


Figure 2-20: Perspective projection of the real point R to the image plane $\bar{\Omega}$ [73].

Chapter 3

Our approach using the Blinn-Phong reflectance with different perspective projections

3.1 Introduction

3.1.1 Our contributions

We briefly explain contributions that our method yields over the previous models in the field.

1. As mentioned in [36], a successful reflectance model for 3D reconstruction of objects should combine two major components: diffuse lobe and specular lobe, because reflectance characteristic of real world surfaces is not the same across the entire surface. As a result, the novel method we propose involves the conceptual advantages of considering perspective projection and non-Lambertian reflectance simultaneously based on the complete Blinn-Phong model composed of both diffuse and also specular lobes [25, 102].
2. As another originality of our work, we consider specular light (showing the ratio of specular reflected light) as well as diffuse light. Furthermore, large values of shininess are imposed in our approach. Combining all these features leads to produce strong specularities in our input images and makes the problem more challenging. We think that the superiority

of our approach to handle these intense specularities which we demonstrate in our experiments is undeniable.

3. We advanced the complete Blinn-Phong model with two various perspective projection approaches and could cope with the arising highly non linear frameworks in both cases, while very few works address perspective projection in their method because of the inherent difficulty of several non linear terms in this kind of projection. Moreover, these perspective approaches will be compared and investigated from different aspects.

4. In order to tackle the problem of applying perspective projection within complicated reflectance equations, we introduce a novel heuristic partial linearization strategy which makes the problem easier to solve, cf. section 3.2.1. This scheme can be used in future researches as a basis for solving the challenging work of combining perspective projection and even more advanced reflectance.

5. Finally, as the CCD camera (a camera with the charge-coupled device sensor) is one of the most important perspective cameras [57], we investigate the effect of modelling CCD camera in our method and the quality of results.

Involving the mentioned model assumptions leads to a concrete PS algorithm as sketched in Fig. 3-1.

Our work extends the approach presented by Khanian et al. [74]. A main point of the latter paper is to study the effect of important task of lightening directions on numerical stability while the presentation is restricted there to one spatial dimension and signal reconstruction. The investigations there has motivated us to consider an indicator for good lighting conditions in 3D reconstruction as presented in the next section.

3.1.2 Related works

Many studies in PS considered non-Lambertian effects as outliers and tried to remove them. Mukaigawa et al. [92] suggested a random sample consensus based approach where only diffuse reflection is selected from among the candidates. Mallick et al. [86] introduced a rotation transformation for transforming the RGB color channel to a SUV color channel with the specular channel S and diffuse channels UV. Then, the specular channel S

is used for removing specularities. Chanki et al. [152] introduced a strategy based on a maximum feasible subsystem approach. In their method, the maximum subset of images satisfying the Lambertian constraint is obtained among the whole set of PS images that include non-Lambertian effects like specularities. A median filtering technique is illustrated by Miyazaki et al. [91] to evade the influence of specular reflections which they considered as outliers. Another method relying on this concept is presented by Tang et al. [131] who proposed a coupled Markov Random Field based on treating the specularities and shadows as noise. Wu et al. [147] considered the 3D recovery problem using a convex optimization technique for separating specularities as deviations from the basic Lambertian assumption in the objective function. Smith et al. [122] used a model-based approach that excludes observations that do not fit the Lambertian image formation model. Hertzmann et al. [60] employed some reference objects which are considered to be of homogeneous material for simplicity, meaning that purely specular or purely diffuse materials are addressed. In some other works more complex appearance models are fitted to estimated data, thereby relying e.g. as in the work of Goldman et al. [54] on the use of a convex combination of a small number of known materials, or as in the paper of Oxholm et al. [98] on a probabilistic formulation for linking geometry and lighting estimation by introducing priors.

Regarding the perspective projection, one of the first works combining this technique with PS is performed by Galo et al. [48]. Their work relies on considering point light sources proximate to the lighted object surface. A perspective PS model is also proposed by Tankus et al. based on the Lambertian reflection [133]. A technically different perspective method for Lambertian PS using hyperbolic Partial Differential Equations (PDEs) is presented by Mecca et al. [90]. Turning to the use of non-Lambertian surface reflectance to account for specular highlights in photometric methods, we may note that the investigation of a shape-from-shading method using the Phong model has been shown to give very reasonable results when employing it within a useful process chain [142]. Therefore it seems apparent that an extension to PS making use of the Blinn-Phong image irradiance equation yields even better results given that PS is proposed to solve the ill-posed problem of shape from shading and advantage of Blinn-Phong model over the Phong model as described before.

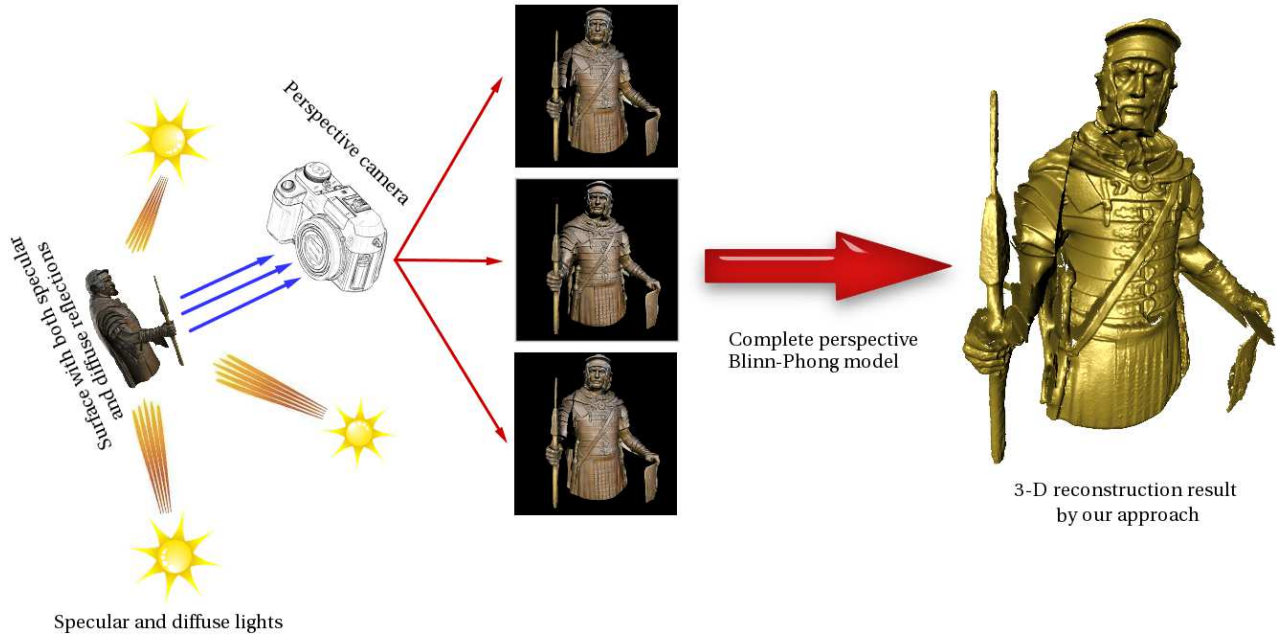


Figure 3-1: Highly specular photometric stereo setup illustrated by a complex synthetic experiment [73]. In real world, surfaces show both specular and diffuse reflections, so considering only diffuse component **or** specular component singly is not enough for real world applications. This surface is illuminated by three non-coplanar light sources with both specular and diffuse lights. Shading due to each light is captured in a perspective CCD camera. As can be seen, considering all stated assumptions, we are able to recover shape with high degree of surface details.

By the combination of the mentioned benefits, we propose a more robust and effectively easier to use method than previous literature, as we detail below. As a side note, since the complete Blinn-Phong model we employ is extensively studied in computer graphics, the surface reflectance in input images as well as expected computational results are potentially easier to interpret than in methods that rely on complex preprocessing steps.

In the following, we deal with the investigation of advancing our method with two different approaches of realizing the perspective projection. The first method is to compute the normal field and then modifying the gradient field based on the perspective projection which is also proposed in [106, 99]. As it manipulates the normal vectors, we refer to this technique as the perspective projection based on the normal field (PPN) method. The second method is to consider a perspective parameterization of photographed object surfaces for getting the gradient field of the surface. We call this approach the perspective projection based on

the surface parameterization (PPS) method.

As we will also consider for experimental comparison a Lambertian perspective PS model, we also recall its construction here. We recall that the Lambertian scene with albedo k_d which is illuminated from directions $L_k = (\alpha_k, \beta_k, \gamma_k)^\top$, where $k = 1, 2, 3$, by corresponding point light sources at infinity, with diffuse intensity l_d , satisfies the following reflectance equation [64]:

$$I_k(x, y) = k_d \left(\frac{L_k \cdot N(x, y)}{\|L_k\| \|N(x, y)\|} \right) l_d \quad (3.1)$$

3.1.3 Modifying normal vectors

The first perspective projection method deals with processing the field of normal vectors $N(x, y) = (n_1(x, y), n_2(x, y), n_3(x, y))^\top$. Once the normal vectors are reconstructed from the orthographic image irradiance equations, the depth map is recovered by giving the following components in to the integrator:

$$p(x, y) = \frac{-n_1(x, y)}{d(x, y)}, \quad q(x, y) = \frac{-n_2(x, y)}{d(x, y)} \quad (3.2)$$

where (p, q) constitute the perspective gradient field for points (x, y) in $\bar{\Omega}$ as the image plane and $d(x, y)$ for a camera with the focal length f is:

$$d(x, y) = xn_1(x, y) + yn_2(x, y) + fn_3(x, y) \quad (3.3)$$

d is the dot product of OP and normal vector $N(x, y)$ where OP is the radial distance of the sensor point $P := (x, y, f)$ to the optical center O when the camera is located at the origin of the coordinate system. In what follows, we show the strategy of this projection in Algorithm 1. The perspective projection realised via projection of the normal vector is denoted by PPN.

Algorithm 1: Transferring orthographic normal field to the perspective gradient field

- 1: Obtain normal field $N(x, y)$ from the orthographic image irradiance equation for $(x, y) \in \bar{\Omega}$ as the image plane
 - 2: Define the distance $OP := (x, y, f)$ as the radial distance of the sensor point P to the optical center O
 - 3: Constitute the parameter of $d(x, y)$ as the dot product of $N(x, y)$ and $OP(x, y)$ as in equation (3.3)
 - 4: Compute the perspective gradient field (p, q) using (3.2)
-

3.1.4 Direct perspective surface parameterization

Another approach to apply the perspective projection is via corresponding surface parameterization, shown in Fig. 2-20. So, in this case, the obtained surface normal (2.34) will be used in image irradiance equation. We recall here the Lambertian perspective image irradiance equation [133], as this will be extended in our model.

In order to remove the dependency of the image irradiance equation on the unknown depth z , it will be substituted by $\nu = \ln(z)$, $z_x = z\nu_x$, $z_y = z\nu_y$, so that we have to apply $z = \exp(\nu)$ to obtain the depth z out of our new unknown ν . This yields:

$$I_k(x, y) = k_d \frac{f\alpha_k\nu_x + f\beta_k\nu_y + \gamma_k(y\nu_y + x\nu_x + 1)}{\sqrt{(f\nu_x)^2 + (f\nu_y)^2 + (y\nu_y + x\nu_x + 1)^2} \|L_k\|} l_d \quad (3.4)$$

A closed form solution for the gradient field is obtained in [133]. For completeness of the presentation, we now recall the main points in its construction. Let us consider three input images (the minimum needed inputs in classic PS). By finding k_d from the first image irradiance equation in (3.4), and replacing it in the second and third image irradiance equation, a linear system of equations $MX = H$ should be solved for obtaining the unknown vector $X = (\nu_x, \nu_y)$:

$$M = \begin{pmatrix} m_1 & m_2 \\ m_3 & m_4 \end{pmatrix}, \quad H = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \quad (3.5)$$

where, we have with $r_i = I_i \|L_i\|$:

$$m_1 = r_2(f\alpha_1 + x\gamma_1) - r_1(f\alpha_2 + x\gamma_2) \quad (3.6)$$

$$m_2 = r_2(f\beta_1 + y\gamma_1) - r_1(f\beta_2 + y\gamma_2) \quad (3.7)$$

$$m_3 = r_3(f\alpha_1 + x\gamma_1) - r_1(f\alpha_3 + x\gamma_3) \quad (3.8)$$

$$m_4 = r_3(f\beta_1 + y\gamma_1) - r_1f(\beta_3 + y\gamma_3) \quad (3.9)$$

$$h_1 = -r_2\gamma_1 + r_1\gamma_2, h_2 = -r_3\gamma_1 + r_1\gamma_3 \quad (3.10)$$

The explicit solutions are:

$$\nu_x = \frac{h_1 m_4 - m_2 h_2}{m_1 m_4 - m_2 m_3}, \nu_y = \frac{m_1 h_2 - h_1 m_3}{m_1 m_4 - m_2 m_3} \quad (3.11)$$

Now, we can obtain the albedo of the surface by plugging the resultant gradient vector for instance into the following equation:

$$k_d = \frac{I_1 \|L_1\| \sqrt{(f\nu_x)^2 + (f\nu_y)^2 + (y\nu_y + x\nu_x + 1)^2}}{l_d \sqrt{(f\alpha_1\nu_x) + (f\beta_1\nu_y) + \gamma_1(y\nu_y + x\nu_x + 1)}} \quad (3.12)$$

3.1.5 Sensitivity of the solution

We try to access the sensitivity of the solution with respect to the lighting directions, which may lead to conditions on the illumination. To this end, the non-singularity condition of the matrix of coefficients M introduced in the previous paragraph should be explored. So, after computing the determinant of M and considering the non-singularity condition $\det M \neq 0$, the non-singularity can be assured in virtually all cases by ensuring that the contributing terms are not zero. This idea leads to the indicator shown in (3.13).

The first three expressions imply the linear independence of light directions and it can be also obtained from the non-singularity condition of the light directions matrix. It should be noted that since PPN deals with obtaining normal vectors from orthographic image

irradiance equation its necessary condition for lights is the non-singularity of the light directions. The other resultant expressions are different and satisfying all of them may not be an easy task. Consequently, the sensitivity of the solution to the lightening in PPS technique can be higher than the PPN approach.

$$\left\{ \begin{array}{l} \beta_1\alpha_3 - \alpha_1\beta_3 \neq 0 \\ \beta_2\alpha_1 - \alpha_2\beta_1 \neq 0 \\ \alpha_2\beta_3 - \beta_2\alpha_3 \neq 0 \\ y\alpha_1\gamma_1 - x\beta_1\gamma_1 \neq 0 \\ x\beta_2\alpha_1 - y\alpha_2\gamma_1 \neq 0 \\ y\alpha_2\gamma_3 - x\beta_2\gamma_3 \neq 0 \\ x\gamma_1\beta_1 - y\gamma_1\alpha_1 \neq 0 \\ y\gamma_1\alpha_3 - x\gamma_1\beta_3 \neq 0 \\ y\gamma_2\alpha_1 - x\gamma_2\beta_1 \neq 0 \\ y\alpha_2\gamma_1 - x\beta_2\gamma_1 \neq 0 \\ x\gamma_2\beta_3 - y\gamma_2\alpha_3 \neq 0 \end{array} \right\} \quad (3.13)$$

3.2 Perspective Blinn-Phong reflectance model

Now we want to develop the Blinn-Phong model by introducing the perspective projection to this reflectance model. To this end, we apply again the two different mentioned perspective approaches. We recall the Blinn-Phong image irradiance equation as:

$$I(x, y) = k_d \left(\frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|} \right) l_d + k_s \left(\frac{H(x, y, z) \cdot N(x, y)}{\|H(x, y, z)\| \|N(x, y)\|} \right)^n l_s \quad (3.14)$$

To develop the perspective Blinn-Phong PS model, we focus on the surface parameterization and plug in the perspective normal (2.34) in (3.14).

Considering k input images for corresponding lighting directions, this yields after some

computation the perspective Blinn-Phong reflectance equations as:

$$I_k(x, y) = k_d l_d \frac{f\nu_x\alpha_k + f\nu_y\beta_k + \gamma_k w}{g\sqrt{(f\nu_x)^2 + (f\nu_y)^2 + (w)^2}} + k_s l_s \left(\frac{f\nu_x D(1) + f\nu_y D(2) + (w)D(3)}{\sqrt{r + (w)^2} \|D\|} \right)^n \quad (3.15)$$

where

$$p := \|(x, y, f)\|, \quad g := \|L_k\| \quad (3.16)$$

$$r := f^2(\nu_x^2 + \nu_y^2) \quad (3.17)$$

$$D := \begin{bmatrix} \alpha_k p + gx \\ \beta_k p + gy \\ \gamma_k p - gf \end{bmatrix} \quad (3.18)$$

$$w = (y\nu_y + x\nu_x + 1) \quad (3.19)$$

3.2.1 Numerical approach

Now, we present the numerical procedure which can be applied for addressing such a highly nonlinear system of equations. Recalling the description of a system of equations as $F(X) = 0$, where $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a given function by the equations from (3.15), we will discuss our solution procedure.

In order to cope with such a nonlinear system of equations, we applied the Levenberg-Marquardt method introduced in [80, 87] as a combination of the Gauß-Newton method and steepest descent direction technique. In this method, if X^k is the point at iteration k , the next iteration can be computed as:

$$X^{k+1} := X^k + d^k, \quad (3.20)$$

$$d^k := -(JF(X^k)JF(X^k)^T + \lambda_k I)^{-1} JF(X^k)F(X^k) \quad (3.21)$$

with $\lambda_k > 0$.

The matrix $JF(X^k)JF(X^k)^T + \lambda_k I$ is positive definite and d^k is well-defined. In addition, this method does not need the conditions such as the invertibility of Jacobian matrix or Hessian matrix or $m = n$.

Partial linearization strategy

Our numerical approach for the PPS method is based on the following formulation which makes the problem much easier to handle by providing a partially linearized system from that highly nonlinear system of equations. Recalling the perspective Blinn-Phong reflectance equations (3.15), and dividing three equations (I_1/I_2 , I_2/I_3 , I_1/I_3 , corresponding to the three used images in our method) leads to a system of equations, with the equations like the following equation as obtained for dividing the first and second images:

- *Integrating two terms in equation (3.15) which results in:*

$$I_k(x, y) = \frac{A(f\nu_x\alpha_k + f\nu_y\beta_k + \gamma_k w)(\sqrt{r + (w)^2}\|D\|)^n + BC^n}{g\sqrt{(f\nu_x)^2 + (f\nu_y)^2 + (w)^2}(\sqrt{r + (w)^2}\|D\|)^n} \quad (3.22)$$

where

$$A := k_d l_d, \quad B := k_s l_s g \left(\sqrt{r + (w)^2} \right) \quad (3.23)$$

$$C := \left(f\nu_x D(1) + f\nu_y D(2) + (w)D(3) \right)^n \quad (3.24)$$

- *Dividing two of irradiance equations for the first and second images*

leading to:

$$I_1(x, y)/I_2(x, y) =$$

$$\frac{\left(A(f\nu_x\alpha_1 + f\nu_y\beta_1 + \gamma_1w)(\sqrt{r + (w)^2}\|K\|)^n + B_1C_1^n \right) F}{\left(A(f\nu_x\alpha_2 + f\nu_y\beta_2 + \gamma_2w)(\sqrt{r + (w)^2}\|S\|)^n + B_2C_2^n \right) G} \quad (3.25)$$

where

$$B_1 := k_s l_s \|L_1\| \left(\sqrt{r + (w)^2} \right) \quad (3.26)$$

$$C_1 := \left(f\nu_x K(1) + f\nu_y K(2) + (w)K(3) \right)^n \quad (3.27)$$

$$\text{and } K := \begin{bmatrix} \alpha_1 p + \|L_1\| x \\ \beta_1 p + \|L_1\| y \\ \gamma_1 p - \|L_1\| f \end{bmatrix} \quad (3.28)$$

$$B_2 := k_s l_s \|L_2\| \left(\sqrt{r + (w)^2} \right) \quad (3.29)$$

$$C_2 := \left(f\nu_x S(1) + f\nu_y S(2) + (w)S(3) \right)^n \quad (3.30)$$

$$\text{and } S := \begin{bmatrix} \alpha_2 p + \|L_2\| x \\ \beta_2 p + \|L_2\| y \\ \gamma_2 p - \|L_2\| f \end{bmatrix} \quad (3.31)$$

$$F := \|L_2\| \left(\|S\| \right)^n \quad (3.32)$$

$$G := \|L_1\| \left(\|K\| \right)^n \quad (3.33)$$

- *Unifying two fractions in (3.25) as follows:*

$$I_1(x, y) \left(A(f\nu_x\alpha_2 + f\nu_y\beta_2 + \gamma_2w)(\sqrt{r + (w)^2}\|S\|)^n \right)$$

$$\begin{aligned}
& + B_2 C_2^n \Big) G - I_2(x, y) \Big(A(f\nu_x \alpha_1 + f\nu_y \beta_1 + \gamma_1 w) \\
& (\sqrt{r + (w)^2} \|K\|)^n + B_1 C_1^n \Big) F = 0
\end{aligned} \tag{3.34}$$

By proceeding the same approach for constituting (I_2/I_3) and (I_1/I_3) , the partially linearized system of equations will be arised. It should be noted that even in this case of existing specularities and in the process of solving the perspective PS system for the Blinn-Phong model (3.15), we will still follow Woodham and make use of only three input images.

Furthermore, as for the case of Lambertian PS, we will also deal with the Blinn-Phong model using the perspective version based on transforming the normal vectors (PPN method), i.e. after orthographic Blinn-Phong PS. Finally, the obtained gradient fields are processed by the Poisson integrator. see e.g. [19] for a recent account of surface normal integration.

3.3 CCD cameras

We will also investigate the modeling of the CCD camera. In the case of CCD cameras, the following projection mapping is used as presented in [57]. The matrix

$$\Gamma = \begin{bmatrix} \psi_x & \xi & \delta_x \\ 0 & \psi_y & \delta_y \\ 0 & 0 & 1 \end{bmatrix} \tag{3.35}$$

contains the intrinsic parameters of the camera, namely the focal length in x - and y - direction equal to $\psi_x = \frac{f}{h_x}$ and $\psi_y = \frac{f}{h_y}$, with the sensor sizes h_x and h_y and the principal point or focal point $(\delta_x, \delta_y)^\top$. The parameter ξ is called skew parameter. Here, we neglect this parameter since it will be zero for most of normal cameras [57]. Using this matrix, we will introduce the following transformation to convert the dimensionless pixel coordinate $X = (x, y, 1)^\top$ to the image coordinate $\chi = (c, d, f)^\top$ as follows:

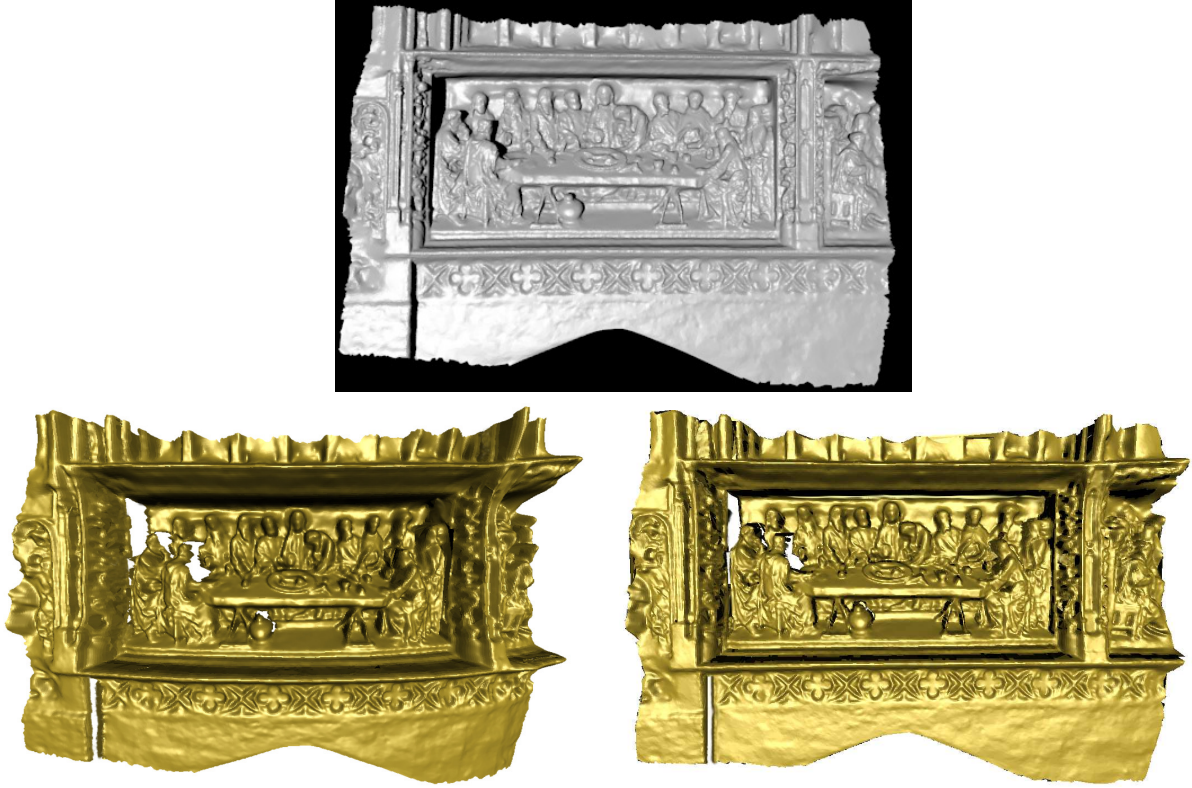


Figure 3-2: Comparison of the surface reconstruction techniques [73]. Top: Input image. Bottom left: Our 3D reconstruction using orthographic projection. Bottom right: Our 3D reconstruction by perspective projection. It can be observed that the perspective approach is able to generate a more compatible result with respect to the original image.

$$X = \Gamma \frac{1}{f} \chi \quad (3.36)$$

By applying the above-mentioned transformation, the following representation for the projected point χ will be obtained:

$$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} h_x(x - \delta_x) \\ h_y(y - \delta_y) \end{bmatrix} \quad (3.37)$$

The effect of this modeling can be potentially interesting, since this information is not always accessible. The above transformation is called *centerizing* in the experiments.

3.4 Experiments

This section describes our experiments performed by the proposed model and approaches. In a first test we confirm the investigation of Tankus et al. [135] that the use of an orthographic camera model may yield apparent distortions in the reconstruction while a perspective model may take the geometry better into account, see the experiment documented in Fig. 3-2. This justifies the use of the perspective camera model. Note that in the figure the object of interest is relatively close to the camera.

In a series of tests we now turn to quantitative evaluations of the proposed computational models. To this end, we consider the set of test images in the next experiments as shown in Fig. 3-4. The Beethoven test images (which depict a real world scene) and the Sphere images are of the size 128×128 . The Stanford Bunny test images have a resolution of 150×120 . Both Bunny and Sphere are rendered using Blender. The 3D model of Stanford Bunny is obtained from the Stanford 3D scanning repository [13]. The 3D model of the face presented in Fig. 3-11 is taken from [126] with the size of 256×256 . For comparing our results, the ground truth depth maps are extracted, and we will make use of the Mean Squared Error (MSE) showing the accuracy.

After considering the mentioned test settings, we demonstrate the applicability of our method at hand of real world medical test images from gastro endoscopy and discuss its superior reconstruction capabilities compared to previous models.

3.4.1 Tests of accuracy

In the first evaluation, we compare the results of two mentioned perspective techniques of PPN and PPS, applied to the specular Sphere in Fig. 3-4 (c) with different values of focal length. MSE results of these 3D reconstructions are shown in Fig. 3-3. While obtained results of described perspective methods for some low values of focal lengths are close to each other, PPN perspective strategy outperforms PPS as the focal length increases. In the second experiment concerned with the Beethoven image set, we investigate the difference between two mentioned perspective approaches on a more complex real world object scene. To this end, we give in Table 3.1 the MSE comparing gray value data of the reprojected and

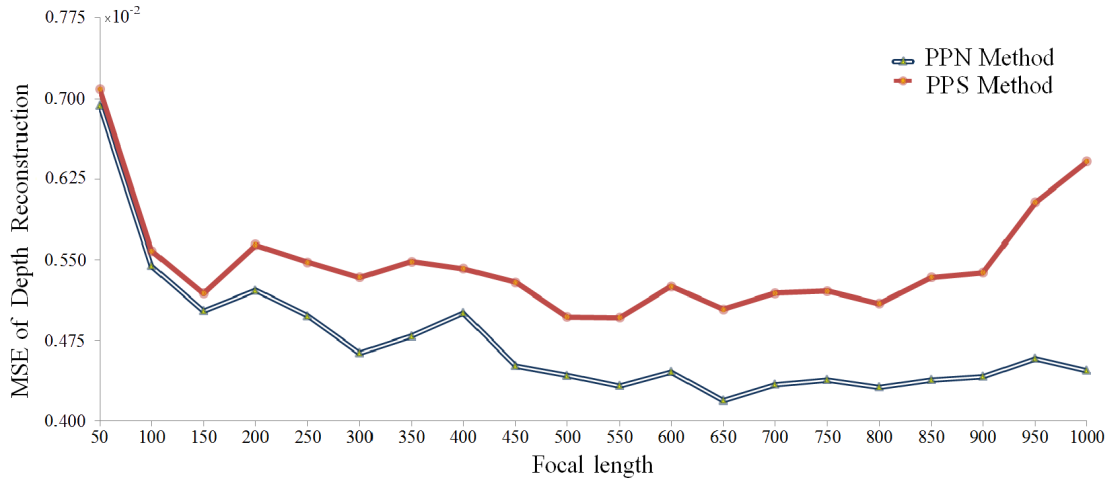


Figure 3-3: Comparing our results with two described perspective methods regarding their depth reconstructions [73]. Considering all mentioned innovations yields to an accurate perspective PS able to handle specularities.

input images. Since in this case the ground truth depth map is not available, we reconstruct the reprojected images by obtaining the gradient fields from the mentioned perspective approaches and replacing them in the Lambertian reflectance equation. It can be deduced from Table 3.1 that reprojecting from PPS method reaches a close accuracy regarding the third input image, while the PPN approach achieves higher accuracy in terms of the first and especially second input image.

As the reprojected images in Fig. 3-5 show, the difference between these methods as given in Table 3.1 can be quite significant. Furthermore, it is indicative of higher sensitivity of the PPS method to the lightening than the PPN approach.

Table 3.1: Comparison between MSE of the reprojected Beethoven images from two described perspective methods of PPN and PPS.

Perspective method	MSE for 1st input	MSE for 2nd input	MSE for 3rd input
PPN method	0.004239	0.003297	0.007535
PPS method	0.008042	0.021409	0.007644

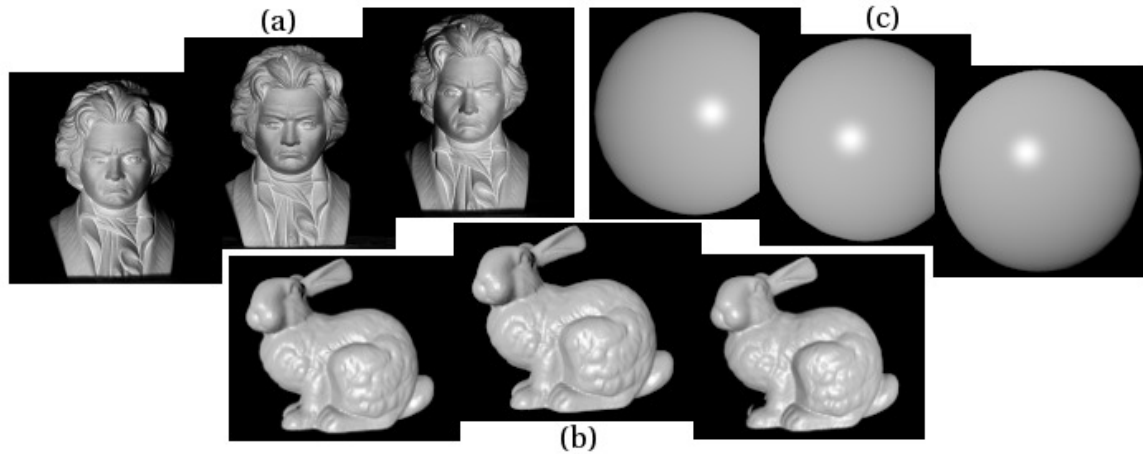


Figure 3-4: Set of three test images used for our 3D reconstruction [73]. (a) Real scene used for reprojecting; (b) and (c) are rendered images used for our 3D reconstruction in presence of specularities.

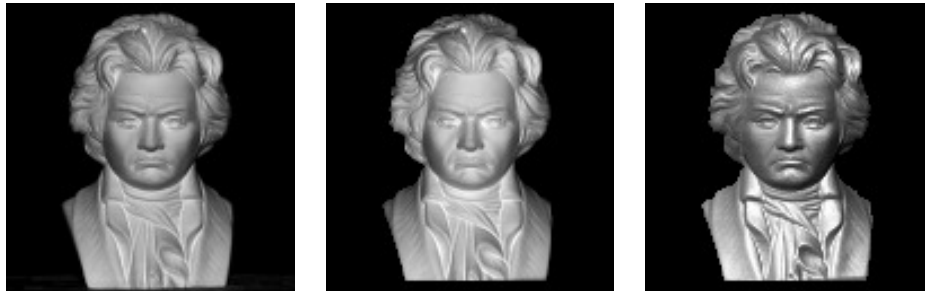


Figure 3-5: An account of reprojected Beethoven images [73]. Left: Second input image for PS. Middle: Reprojected second image obtained from PPN method. Right: Reprojected second image using the PPS technique.

3.4.2 Slant and tilt

Two well-known indicators of 3D information are slant and tilt [96]. Basically, there are some individual neurons in our brain’s caudal intraparietal area (a critical neural locus for encoding 3D information of objects) which are responsible to encode slant and tilt of the surfaces [109, 125]. Here, we performed some other experiments to probe the behavior of both proposed perspective techniques for slant and tilt parameters which can be considered as the perceptual properties of the surface [112, 124]. Slant is defined as the angle between the surface normal and the line of sight, while the tilt angle determines the orientation of the surface normal projection on the fronto-parallel plane (the perpendicular plane to the viewing direction which is also called image plane) shown in Fig. 3-6. These parameters

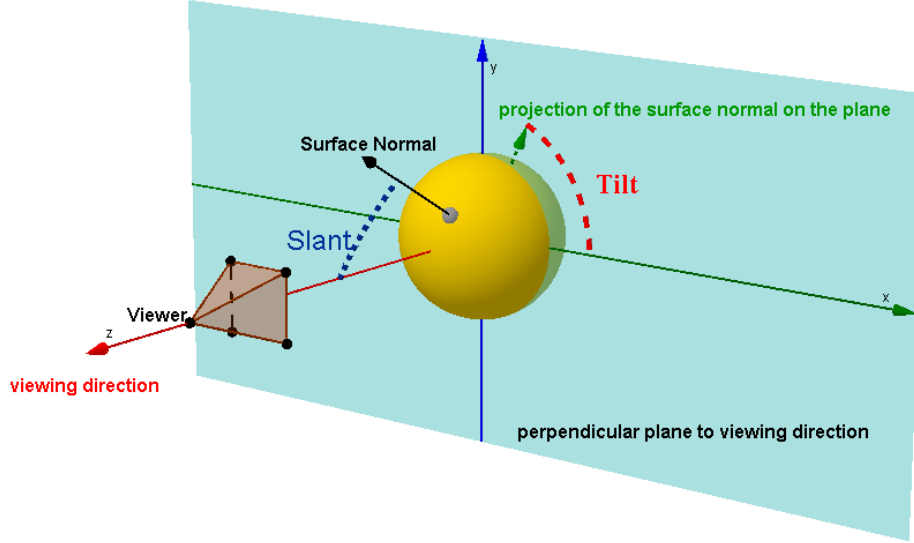


Figure 3-6: Two angular variables of slant and tilt [73] as the important parameters used for encoding the 3D perceptions in our brain [109, 125]. The tilt of a surface corresponds to the direction of largest variation in perceived distance and slant varies with the magnitude of the gradient according to (3.38).

are obtained for the case of two illustrated perspective projection methods.

Since in perspective projection we deal with the surface gradient field (p, q) , slant and tilt formulas will be defined as follows [124]:

$$\forall (x, y) \in \bar{\Omega} :$$

$$\sigma(x, y) = \arctan(\sqrt{p(x, y)^2 + q(x, y)^2}) \quad (3.38)$$

$$\tau(x, y) = \arctan \frac{q(x, y)}{p(x, y)} \quad (3.39)$$

In PPN, the gradient field is computed using equation (3.2), while in PPS it is obtained by:

$$p := z\nu_x, \quad q := z\nu_y \quad (3.40)$$

Variations of the slant and tilt on a unit sphere are shown in Fig. 3-7. As can be seen, the slant parameter changes in the interval of $[0^\circ, 90^\circ]$, whereas tilt varies in $[0^\circ, 180^\circ]$.

To conduct a fair comparison between PPN and PPS influence on 3D information, we gained tilt and slant angles of a unit sphere under those two projections, calculated the

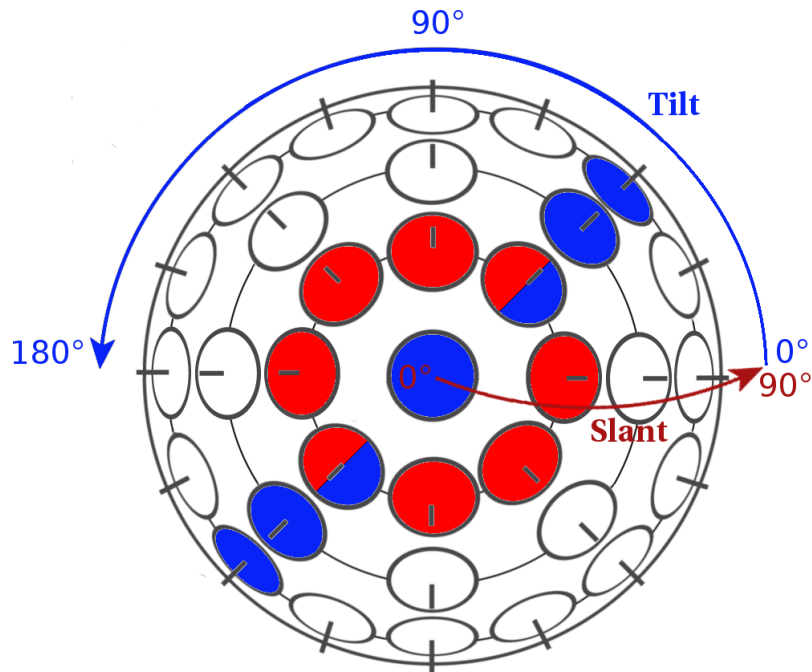


Figure 3-7: Slant (red) and tilt (blue) components of surface orientation on a unit sphere [73]. Red circles (including two bi-colors) represent an area with equal slant angles and blue circles (including two bi-colors) indicate an area with equal tilt angles on the sphere.

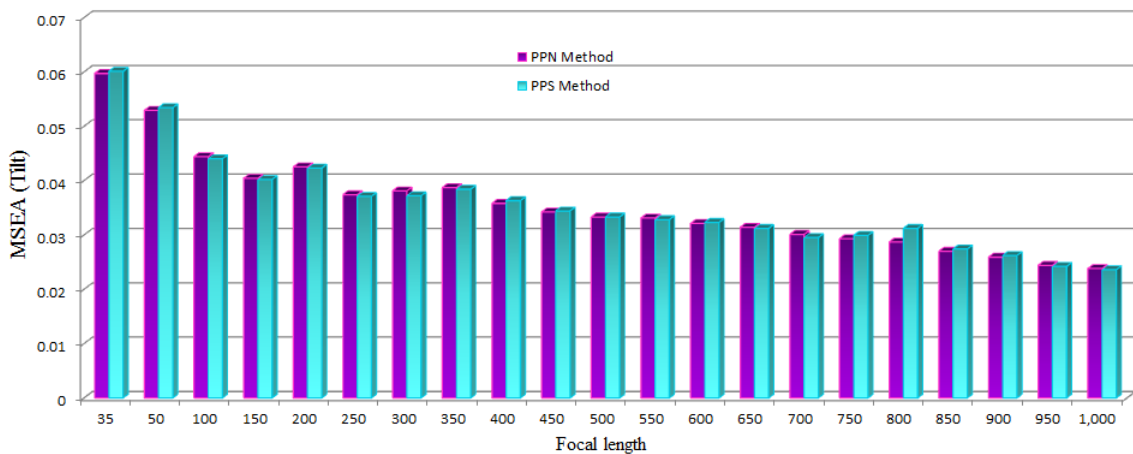


Figure 3-8: Mean square error of tilt angle (MSEA in degrees) for two mentioned perspective techniques applied on the sphere against different focal lengths [73]. As can be seen, the error in estimating tilt angle is decreasing by increasing focal length.

Mean Square Error of obtained Angles (MSEA) and compared them in Fig. 3-8 and Fig. 3-9, respectively.

It is important to note that any deviation in tilt reconstruction leads to false perception of

the *object orientation* which is not detrimental for the perception of a symmetric 3D shape such as a sphere, whilst any failure in slant value reconstruction yields to loss of the object *curvature understanding*. In other words, the geometrical interpretation of an erroneous slant estimation is the distorted curvature of the reconstructed surface.

As depicted in Fig. 3-8 and Fig. 3-9, the gradual increase of the focal length diminishes the estimation error with respect to both tilt and slant angles.

A quick glance on the first chart (Fig. 3-8) reveals that the range of tilt values for two perspective projections remain close together confirming that direction estimation accuracy (tilt impression) is not a discriminative criterion to measure the difference between these two projections in the case of a spherical surface reconstruction.

In contrast, curvature estimation accuracy (slant impression) in Fig. 3-9, conveys the more successful performance of the PPN approach to preserve curvature properties of the surface during 3D reconstruction process.

Our conclusion is in accordance with some previous studies on the human visual system [137, 29], where it turns out perspective characteristics of the human visual system is the main source of slant perception.

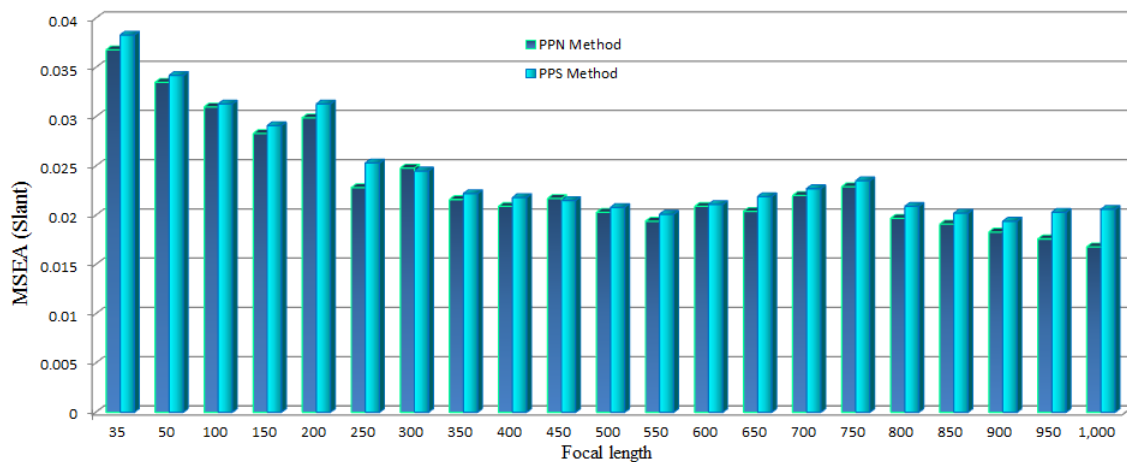


Figure 3-9: Mean square error of slant angle (MSEA in degrees) for two mentioned perspective techniques applied on the sphere against different focal lengths [73]. The trend of error is descending with respect to gradual increasing of focal length. This amount of decreasing can be seen in PPN perspective projection more than PPS perspective technique.

Table 3.2: MSE of the reconstructed depth from images with specularities by two perspective methods of PPN and PPS. As it is clear, we consider 3D reconstruction in the presence of both diffuse and specular reflection simultaneously from the surface which leads to involving both k_d and k_s and applying complete Blinn-Phong model. In addition, we applied both diffuse and specular light. Finally, we extended our model to different perspective projection techniques.

Reconstruction Error	k_d	k_s	l_d	l_s	shininess	centerizing	no centerizing
MSE of PPN (Bunny)	0.6	0.4	1.2	1.2	50	0.006355	0.042082
MSE of PPS (Bunny)	0.6	0.4	1.2	1.2	50	0.012318	0.011318
MSE of PPN (Sphere)	0.5	0.5	1.2	1.2	150	0.008264	0.022568
MSE of PPS (Sphere)	0.5	0.5	1.2	1.2	150	0.008431	0.007716

3.4.3 Perspective methods and CCD camera model

Table 3.2 and Fig. 3-10 present the results of our 3D reconstructions for highly specular input images as shown in Fig. 3-4 (b) and Fig. 3-4 (c), respectively. In order to produce such images, we set non-zero intensities for diffuse and also specular light. Furthermore, the objects include both diffuse and specular reflections. Values of MSE for 3D reconstruction show the high accuracy of our depth reconstructions by applying the complete Blinn-Phong model which is accompanied by two presented perspective schemes. On the other hand, while results of the recovered depth map for the sphere are close to some extent, the outcome of the computed depth map for Bunny based on the PPN method obtains higher accuracy. However, the table also illustrates the higher sensitivity of the PPN perspective scheme to centerizing transformation than the PPS perspective method. Finally, we compare our approach with the Lambertian model which is the most common model applied in PS and also the method presented by Mecca [89]. Last row in Fig. 3-10 shows the outcome of applying the Lambertian model. The deviation from faithful reconstruction over the specular area of the surface can be seen clearly. The comparison between our approach and [89] is also shown in Fig. 3-11. As already indicated, our method applies complete perspective Blinn-Phong model on three images including both diffuse and specular reflections and lights, while the method in [89] uses the specular term in Blinn-Phong model to handle four purely specular images. The excellent result of the proposed method presented in Fig. 3-11 (b) over the high value of specularity with the absence of any de-

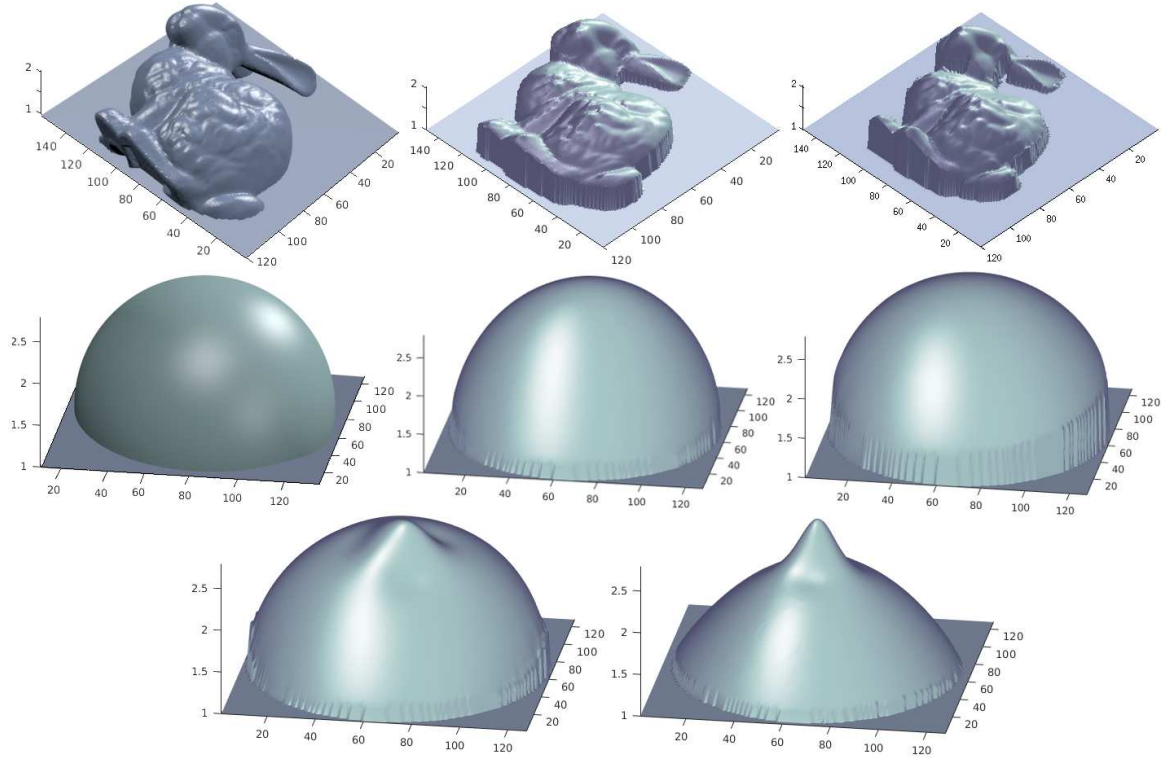


Figure 3-10: First and second row: Left: Groud truth. Middle: depth reconstruction from complete Blinn-Phong model with PPN approach. Right: depth reconstruction from complete Blinn-Phong model with PPS approach. These results turn out the proficiency of the proposed method for appealing reconstruction of the images including strong specularities. In addition, PPN approach achieves more faithful reconstructions. Last row: Depth reconstruction from Lambertian model in the presence of specularity accompanied by different perspective projection. Left: PPN approach. Right: PPS method. As it can be seen, the Lambertian model is not able to provide a faithful reconstruction for the specular surface [73].

viation or artifact shows that the proposed method outperforms state-of-the-art approaches such as in [89]. The MSE values of 3D reconstruction associated with experiments in Fig. 3-11 are also illustrated in Table 3.3.

3.4.4 Tests of applicability on real world test images

This section describes experiments conducted by the proposed approach on realistic images. We first turn to some real world medical test images. It should be noticed that we

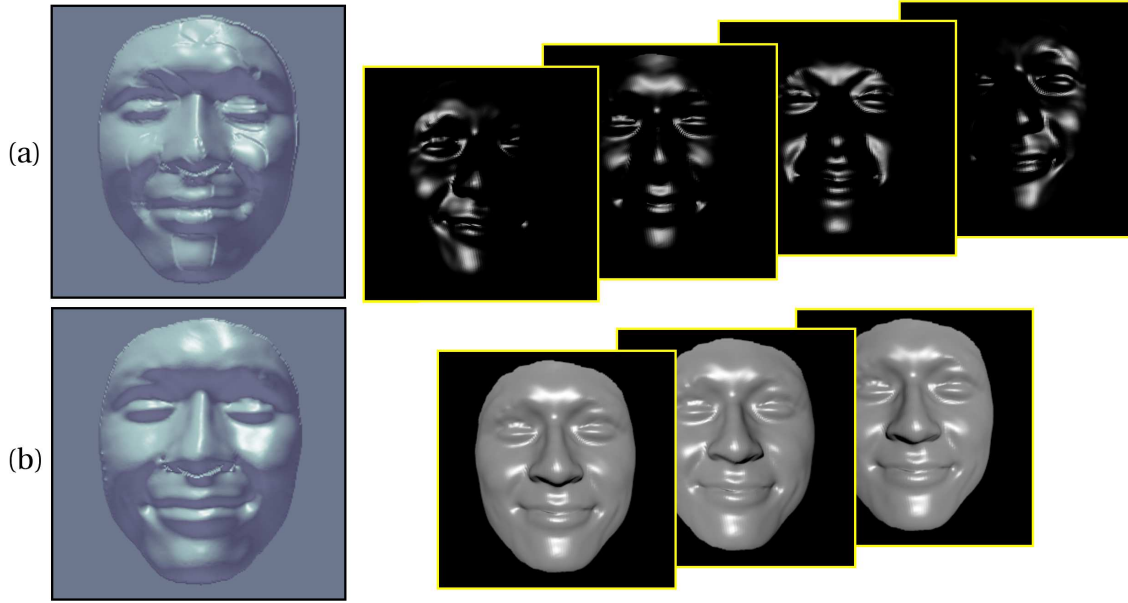


Figure 3-11: First row: Four purely specular input images as applied in purely specular model of [89] and the 3D reconstruction of the [89] approach which shows deviations especially around the highly specular areas. Second row: Three ordinary input images including both diffuse and specular components as the input of our method and Our 3D reconstruction. Note that our method does not need the decomposition of the input images into purely diffuse and purely specular components which is a very difficult task even for synthetic images [73].

may also call these images realistic because we did not benefit from a controlled setup or additional laboratory facilities. We used just the images that are available as in any kind of medical (or many other real world) experiments. Let us note that experiments with endoscopic images are well known to yield a challenging test and they are widely accepted for indicating possible medical applications of photometric approaches, see e.g. [134, 136]. As for our work, the usefulness of computational results for the indicated, concrete medical

Table 3.3: MSE of the reconstructed depth from images with high specularities shown in Figure 3-11.

Depth reconstruction approach	k_d	k_s	l_d	l_s	shininess	MSE
Proposed method	0.3	0.7	1.2	1.2	50	0.004019
Mecca [89]	0	0.7	0	1.2	50	0.056586

application is confirmed via collaboration with specialized medical doctors.¹ We have performed trials on endoscopic images in which existence of high specularities is unavoidable. Input images are presented in Fig. 3-12 (a) and Fig. 3-12 (b) which are endoscopies of the upper gastrointestinal system. Their 3D reconstructions are represented in Fig. 3-13 and Fig. 3-14.

Similar to all the previous experiments only three input images are used. All outputs are displayed with an identical viewpoint enabling their visual comparison. The first column in Fig. 3-13 is indicative of the deviation in the Lambertian result. As it is visible in the cropped region in Fig. 3-12 (a), shown in the rectangular part, the beginning and end points of all three folds (marked by A, B and C) should be at about the same level, instead a drastic deviation toward downside is showing up at the left side of the surface (cf. Fig. 3-13 (a)) in results obtained by applying Lambertian reflectance model as also indicated by the blue area in the corresponding depth map shown in Fig. 3-13 (d).

However, this deviation is rectified by applying the complete Blinn-Phong model accompanied by PPS as can be seen in the second column of the Fig. 3-13 and also entirely corrected using this model with PPN approach represented in the third column in Fig. 3-13. Furthermore, three folds of the surface are reconstructed very well in the Blinn-Phong outcomes (second and third columns of Fig. 3-13). This obviously desirable complete reconstruction of those folds cannot be seen in the Lambertian output.

Finally, as also the color alteration (in depth maps) represented in the second row of Fig. 3-13 shows, high frequency details are recovered as well in the Blinn-Phong outputs especially in the case of PPN approach.

These reconstruction aspects are again clearly observable in another endoscopy image depth reconstruction in Fig. 3-14 which are the depth resultings from inputs as in Fig. 3-12 (b). Once more, a deviation from the desirable output shape appears in the Lambertian outcome especially in the left corner side (Fig. 3-14 (a)). This part of the surface, which is marked by (C) in the input and 3D resulting images, has a cavity toward the upside in reality, which is reconstructed well by the Blinn-Phong outputs in contrast to the

¹We mention as a reference the collaboration with Dr. Mohammad Karami H. (Dr.mokaho@skums.ac.ir) who is a gastroenterologist and internal medicine specialist at Shahrekord University of Medical Science (Iran). The input endoscopy images are also his courtesy.

Lambertian result. The Lambertian model apparently provides a reconstruction completely on the opposite side for this region of the original surface.

Let us pay attention also to the second row in Fig. 3-14 displaying depth maps. A curved line of the upper corrugated region (A) is obtained in the right corner of the Blinn-Phong outputs (Fig. 3-14 (e) and Fig. 3-14 (f)), whereas this region is just a straight line in the right corner of the Lambertian outcome (Fig. 3-14 (d)). The height of corrugated regions are obviously more faithfully reconstructed in the Blinn-Phong results compared to the Lambertian one.

Last but not least, it is worth to mention that the viewing angle of the endoscopy cameras is very tight. Using cropped parts of those images in our experiments makes this experiment a highly challenging task of 3D reconstruction. The success of our approach to reconstruct such a tiny range of the depth values without any knowledge about photographic conditions reveals the capability of our proposed method in challenging real world applications.

In another test with real world input images, we compared our method with the approach used in [133] by making use of the input images depicted as Fig. 2 (a), Fig. 2 (b) and Fig. 2 (c) in [133]. The surface is a plastic mannequin head, and the plastic material itself shows specularities. It is well-known in computer graphics that plastic is a material that can be readily rendered by using the Blinn-Phong model [101].

The depth reconstructions obtained by our technique and method of Tankus for those real world images are presented in Fig. 3-15 and Fig. 3-16. Once again, the deviation from a natural shape in the Lambertian result of [133] can be clearly observed in the output in Fig. 3-15 (b) shown in an identical view with our result in Fig. 3-15 (a). In addition, let us note that the output of the Blinn-Phong model is very clear and smooth, also at highlights. The inhomogeneous recovery of the shape when using the Lambertian model is cropped at some regions such as chin and tip of the nose cf. Fig. 3-15 (c), where we had to turn the Lambertian result to show these regions. The curved line appearing in the chin and the sharp point at the nose in the Lambertian reconstruction are also visible in [133]. Moreover, as proposed in [133], they could not process eyes in images, due to their specularities, while we succeeded in recovering the faithful 3D shape even with eyes using the complete Blinn-Phong model as presented in Fig. 3-16.

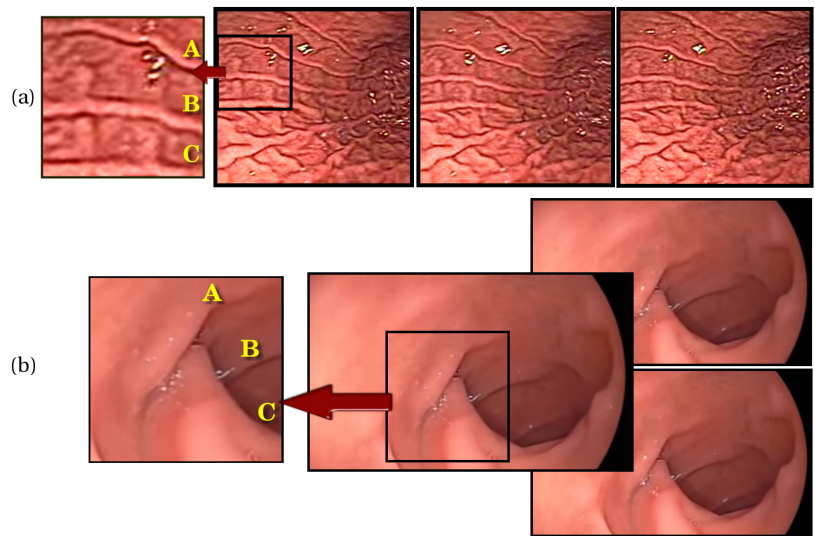


Figure 3-12: Test images with high specularity used in realistic real world scenario [73]. These images are produced in an endoscopy experiment. So, they do not benefit from any laboratory facilities or confine to the controlled setup conditions.

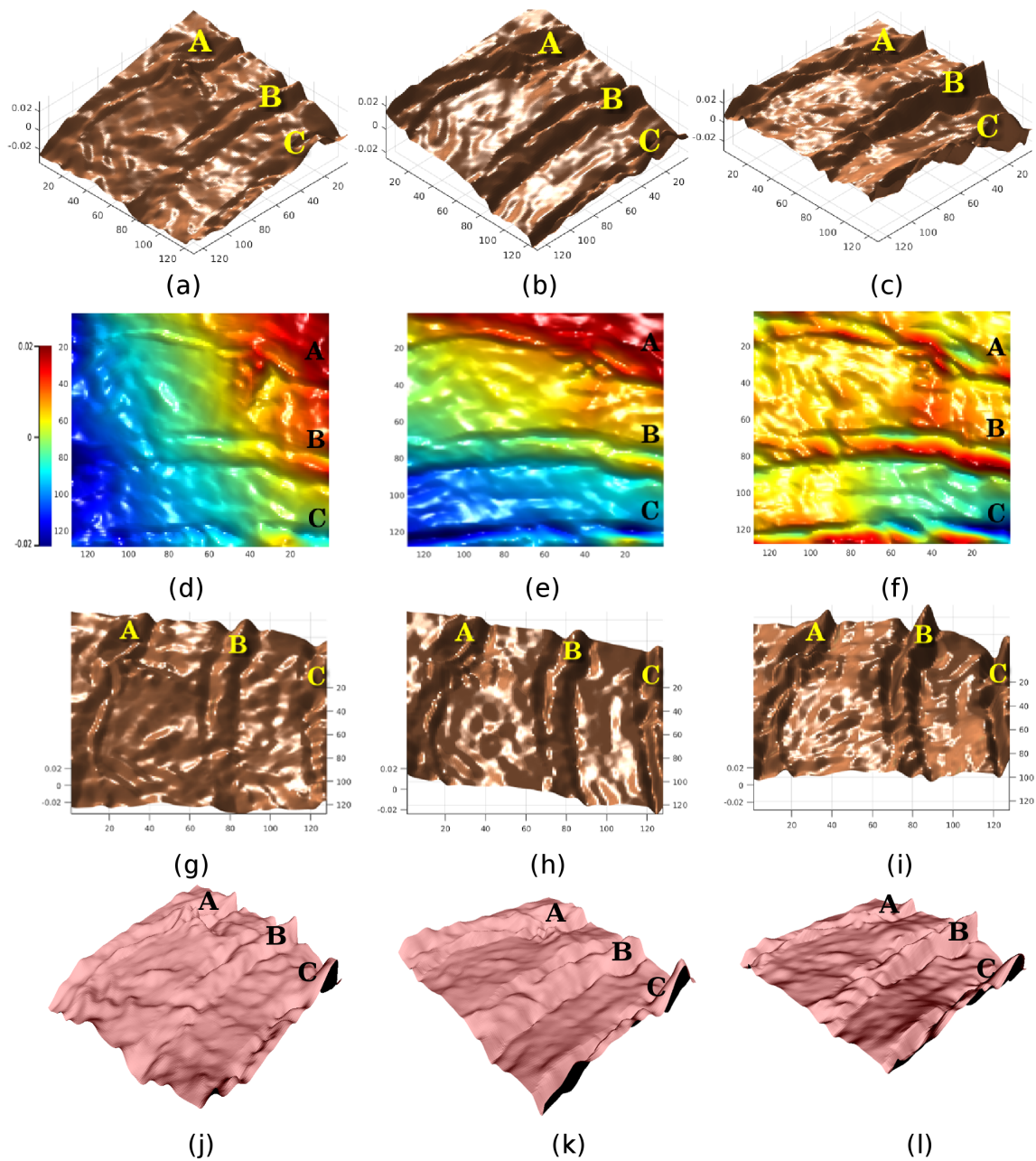


Figure 3-13: Depth reconstruction from real world endoscopy images: (first column) results of Lambertian model, (second column) results of the first proposed method (complete Blinn-Phong using PPS) and (third column) results of second proposed approach (complete Blinn-Phong model using PPN). All images are shown from identical view to show the differences. The deviation in the Lambertian results can be clearly seen, while the results of our approach provide faithful 3D reconstruction without any deviation and also with a high amount of details [73].

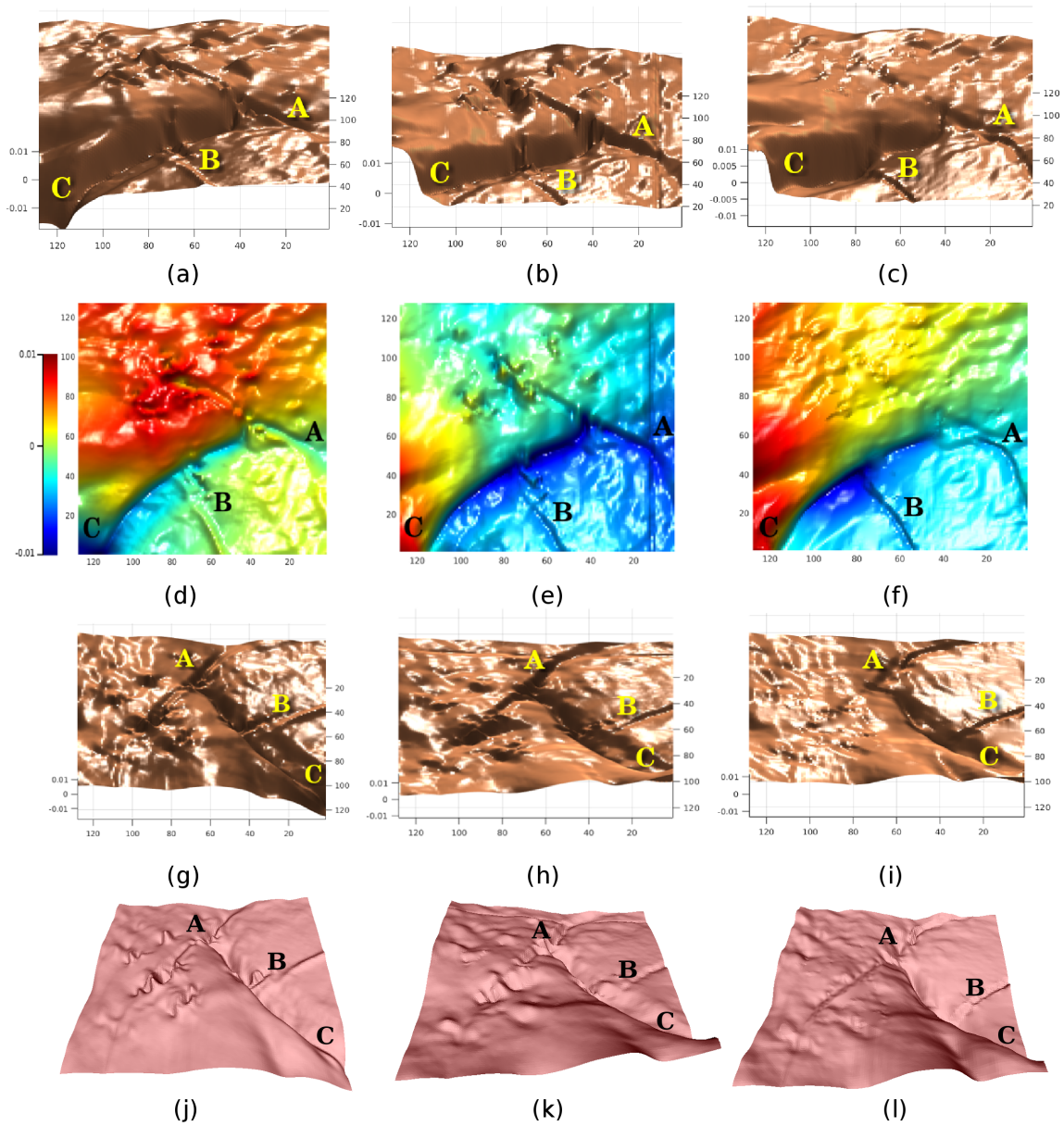


Figure 3-14: Depth reconstruction from real world endoscopy images: (first column) results of Lambertian model, (second column) results of the first proposed method (complete Blinn-Phong using PPS) and (third column) results of second proposed approach (complete Blinn-Phong model using PPN). All images are shown from identical view to show the differences. Once more, the deviation in Lambertian outcomes is clear, whereas our approach provides a trustable 3D reconstruction without any deviation [73].

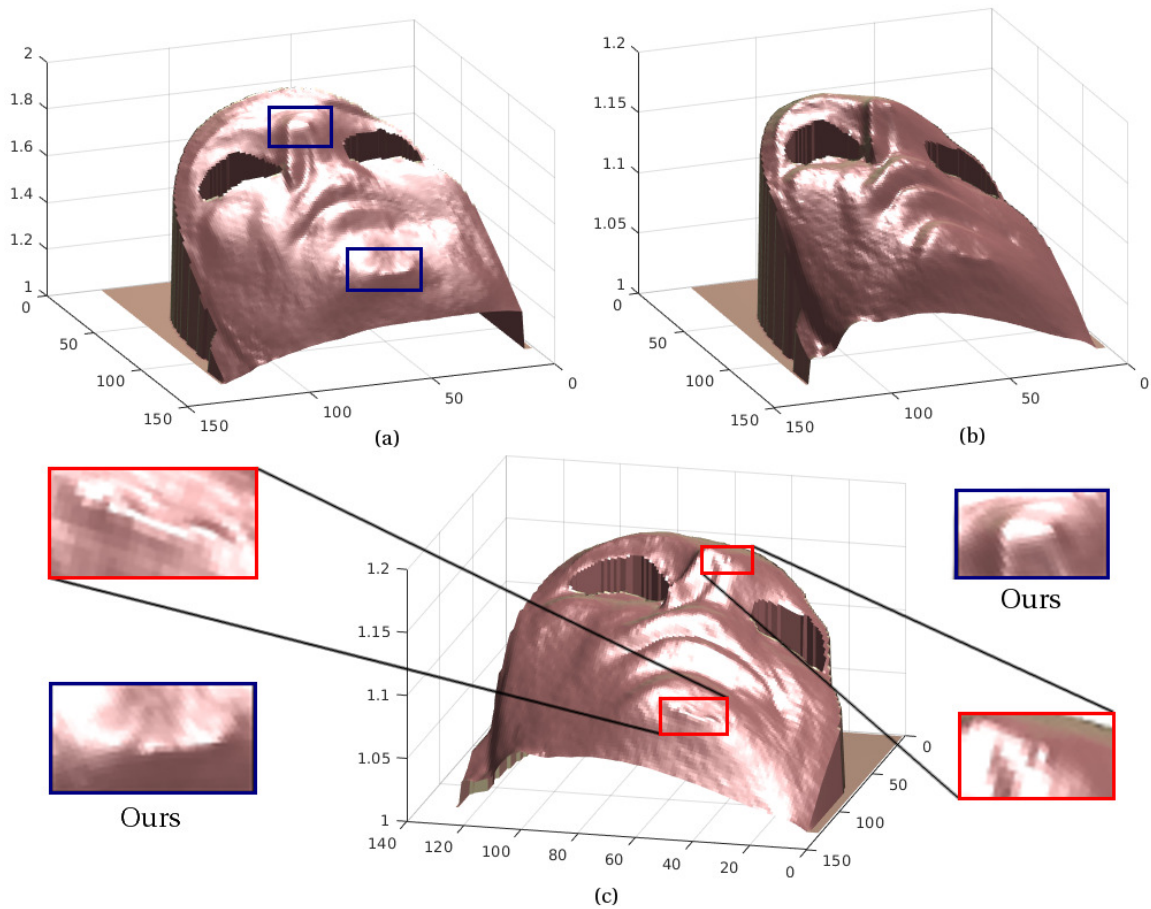


Figure 3-15: Depth reconstructions from real world images: (a) Results of our proposed method using complete Blinn-Phong model, (b), (c) results of [133]. Both images are shown from identical view to represent the differences. We have also cropped some parts of our results and shown them together with the same cropped area of outcomes of [133] in (c). As it is clear, our approach shows significant superiority over [133] in terms of advantages such as smoothness over the rough output of [133], reconstruction success in specularities and absence of deviation from natural symmetric shape [73].

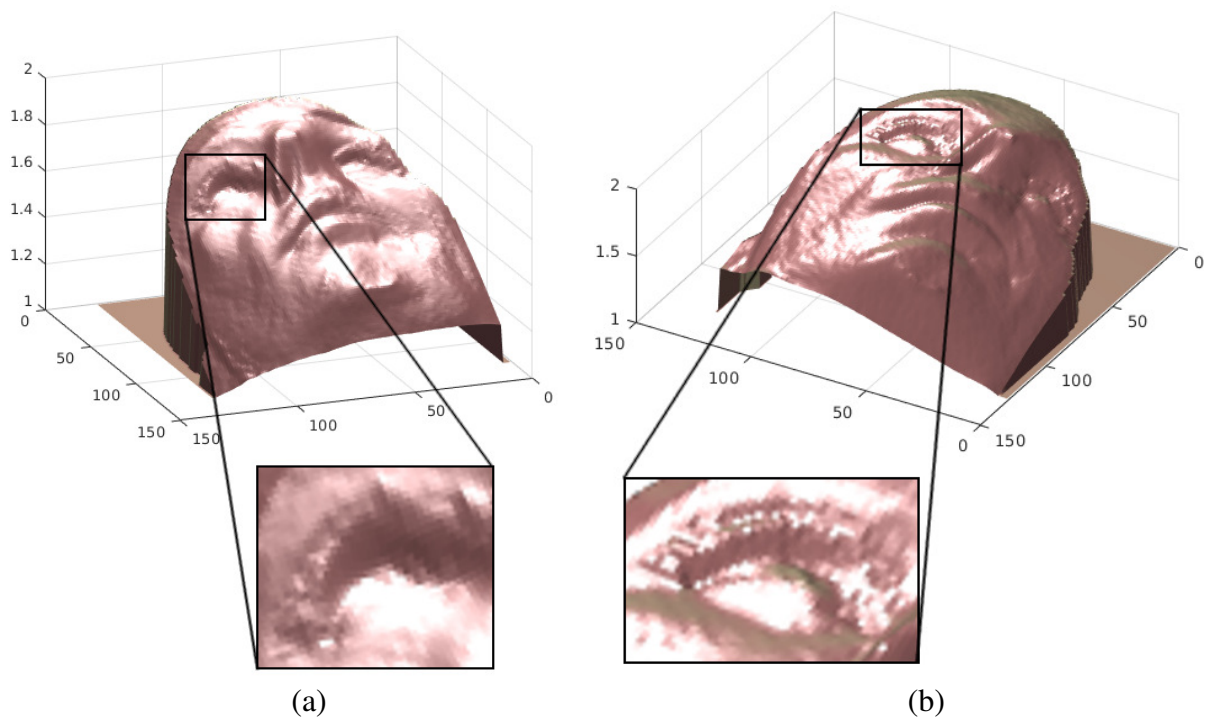


Figure 3-16: Depth reconstructions from real world images: (a) results of our proposed method using complete Blinn-Phong model, (b) results of [133]. As it is mentioned in [133], they could not obtain the reconstruction in the presence of eyes (due to the specularities) unlike our approach which provides faithful results even with including eyes [73].

3.5 Summary and conclusion

A new framework in PS considering the complete perspective Blinn-Phong reflectance including strong specular highlights is presented. The advantages of our method over state-of-the-art PS methods and also the Lambertian model are proved via a variety of experiments. The model includes a perspective camera projection. Furthermore, two different techniques applied in perspective projection are evaluated. In addition, we have also evaluated the modelling of CCD camera. All results are obtained using a minimum necessary number of input images, which is an aspect of practical relevance in different applications and makes PS an interesting technique for close to real-time reconstruction, where a minimal set of images is required. We have demonstrated experimentally also the merits of our PS model for possible challenging real world applications, where we recover the surface with high degree of details. Let us also comment that our computational times are very reasonable i.e. in the order of a few seconds in all experiments.

Concerning possible limitations, as with all the possible approaches that rely on a parametric representation of surface reflectance, the corresponding additional parameters in the reflectance function have to be fixed. This issue may provide challenging numerical aspects in the optimization. Also, while the Blinn-Phong model gives already reasonable results as we demonstrated, other more sophisticated reflectance models may be adequate for handling highly complicated surfaces, which may be a possible issue of future research.

Chapter 4

Real-time 3D shape of micro-details

4.1 Introduction

4.1.1 Our contributions

In this chapter, we provide the following innovations:

1. We proposed an accurate hybrid **real-time** 3D reconstruction approach which merges the benefits of perspective projection and non-Lambertian reflectance model motivated by Cook-Torrance reflectance [37] to handle complex reflecting characteristics. To the best of our knowledge, this is the first work for the real-time PS which also regarded the application of perspective Cook-Torrance model [37].
2. Introducing a **Recurrent Optimization Network (RON)** which provides a robust PS framework to obtain a highly-detailed 3D shape from even **micro-prints** texts where their visibility is hard for the human visual system as can be seen in Fig. 4-1. In this figure, we could reconstruct 3D shape of two regions of **A** and **B** with high details, while these regions are very tiny microtext. The information hidden in these areas became completely clear by our approach. As a result, our technique can be used in different applied areas.
3. Since the performance of many optimization techniques is highly-dependent on the initialization process, we provided a **Dijkstra Gaussian Mean Curvature (DGMC)** technique to find anchor points which can be applied as a key point to offer more effective optimization approach.

4. Furthermore, we extended our proposed approach to deal with the realistic lightening situation using spatially dependent lighting.
5. We show that our approach is able to provide the real-time 3D shape of objects as captured by an ordinary mobile phone camera without employing controlled laboratory conditions (cf. Fig. 4-2). So, our PS process is conducted by the consumer-level equipment where specialized equipment and professional laboratory settings are not required. Our instruments are easily available for general consumers as should be expected for an ideal solution. Consequently, our method can be readily employed in real-world situations. We demonstrate that this attempt provides the basis for a potentially useful 3D reconstruction technique. This improvement breaks inhibitor limitations of a controlled setup and the necessity of working with specific scenes (e.g. Lambertian materials) and turns mobile phone into powerful 3D shape reconstruction tools.
6. Finally, our approach also provides non-uniform colorful albedo from images with diverse color intensities as shown in Fig. 4-12.

All these efforts greatly advance the applicability of photometric stereo to various applications, especially for reconstructing 3D surfaces from very tiny structures like micro-prints.

4.1.2 Related works

In order to circumvent difficulties of specularities, commercial instruments for 3D reconstruction use sophisticated techniques such as white light interferometry or scanning focal microscopy, while these laboratory-based devices tend to be large, slow, or at the expense of \$100,000 or more.

A preprocessing technique is presented by Yang *et al.* [149] to remove specularities. A maximum-likelihood estimation [141] and an expectation maximization [148] are other approaches based on assuming specularities as the outliers. Zickler *et al.* [156] provides an analyze to a subspace which is free from highlights for image transformation. Ikehata *et al.* [68] suggested a regression procedure that deals with the specular component as a sparse error of an underlying purely diffuse reflectance equation. An inpainting technique for the

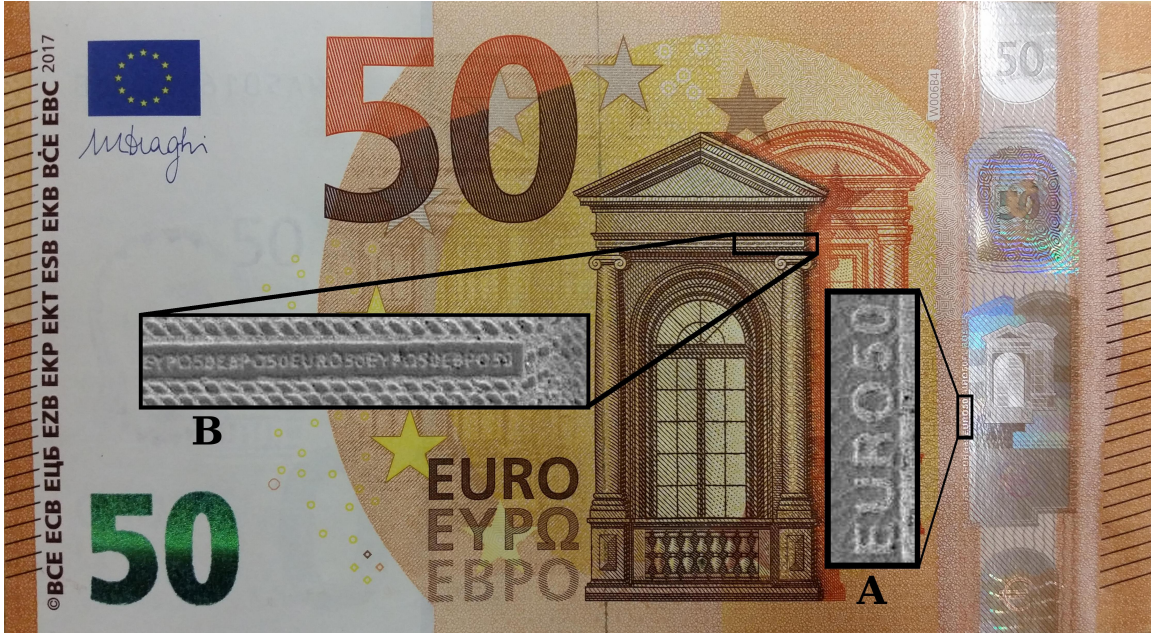


Figure 4-1: A successful 3D reconstruction from very fine details of micro-prints on a 50 euro banknote shown in gray color rectangular parts of (A) and (B) [73]. As can be seen, these details are even invisible to human eyes, while our **real-time** 3D reconstruction from these regions is able to reveal hidden information and features in high amount of details. This information recovery can be used in various applied areas such as detecting security items on financial documents for fraud detection and also the quality control of any industrial productions that include delicate details such as printed circuits.

highlights correction is presented by Tan *et al.* [105] using the color of illuminant and the diffuse component. In [130] and [116] the pixel chromaticity direct analysis and specular-free image are used respectively to specular correction in which the diffuse component is adjusted to a criterion. This criterion is an indicative of smooth transition of color between diffuse and specular regions and evaluated using a nonlinear shifting of specular pixels intensity and chromaticity. Johnson *et al.* [69] presented an elastomer using a sensor skin constructed with a metal-flake pigment which is attached to the surface to change its BRDF. However, in all these works, the shape information that is evidently contained in non-Lambertian effects (e.g. highlights) is discarded.

4.1.3 Lightening sensitivity analysis

In addition to the directional light, we also equipped our perspective Cook-Torrance model with spatially varying lightening. To this aim, a three dimensional vector field of the light



Figure 4-2: A complex real scene including several objects with different sizes captured by a mobile phone with a simple setup which is used for our 3D shape reconstruction [73].

L should be imposed as in (2.31) instead of a directional light vector.

4.1.4 Optimization process

In cases where problems seem hard to solve, different sciences get in different ways to solve the problem. Researchers know that finding solutions to problems can be faster by modeling them in the virtual world and finding the solution in this way can be possible. Therefore, it helped optimization methods to be applied up to now even for cutting edge technologies and fields. The aim of optimization is maximizing desired output as well as minimizing the consumed time. In the nature of optimization, we deal with determining the value of functions whose constraints and limits are clear. In general, problems addressed in optimization techniques can be divided into two types of linear and nonlinear.

Naturally, nonlinear problems are harder to solve and require more complicated optimization algorithms to get the result. This motivated us to devise a RON framework for our approach.

4.1.5 Different optimization algorithms

We recall the Cook-Torrance reflectance designed to cope with the problem of modeling reflections of realistic materials [37]. The Cook-Torrance image irradiance equation is:

$$\begin{aligned}
 I(x, y) = & \underbrace{k_d \frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|}}_{diffuse\ term} l_d \\
 & + \underbrace{\frac{G(x, y, z) D(x, y, z) F(x, y, z)}{4 \frac{V(x, y, z) \cdot N(x, y)}{\|V(x, y, z)\| \|N(x, y)\|} \frac{L \cdot N(x, y)}{\|L\| \|N(x, y)\|}}}_{specular\ term}
 \end{aligned} \tag{4.1}$$

Using equation (4.1) for three input images (as the minimum number of input images in PS), we constitute a system of equation $F(X) = 0$, where $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined as $F(X) = [f_h]$, $h = 1, 2, 3$ and formed by equations given by (4.1) corresponding to each input image $I_h(x, y)$, $h = 1, 2, 3$ as:

$$f_h = I_h(x, y) - (diffuse\ term) - (specular\ term) = 0 \tag{4.2}$$

To estimate the normal field, we should solve the optimization problem as follows:

$$X^* = argmin \frac{1}{2} \|F(X)\|^2 = \frac{1}{2} \sum_{h=1}^3 f_h^2$$

In optimization process, we update the solution using an iterative procedure:

$$X^{k+1} := X^k + d^k, \tag{4.3}$$

Here, d^k is the search direction in iteration k . The main difference between various optimization approaches is caused by the definition of d^k .

4.1.6 Quasi-Newton with BFGS updating

In this section, we investigate various procedures that we applied for the nonlinear system of equations described by $F(x) = 0$. Among many iterative methods, Newton's method is

one of the popular approaches. But this method has some drawbacks which can be removed by its modified versions. Here, we used the updated formula known as Quasi-Newton with BFGS suggested independently by [30, 46, 53, 115], which as explained in [104], has much better performance than other modified versions of Newton's method and it has proven to show good performance even for non-smooth optimizations [81]. In this approach, the Quasi-Newton condition (secant condition) is used to find the search direction d^k :

$$d^k := -(B^{-1})^k \nabla \Psi(X^k) \quad (4.4)$$

where B and $J(X)$ are the Hessian matrix approximation and the Jacobian matrix of $F(X)$, respectively. The following expression is known as Quasi-Newton condition or the BFGS formula for updating B :

$$B^{k+1} = B(X^k) + \frac{Y^k(Y^k)^T}{(\Theta^k)^T Y^k} - \frac{U^k(U^k)^T}{(\Theta^k)^T U^k} \quad (4.5)$$

with

$$Y^k = \nabla \Psi(X^{k+1}) - \nabla \Psi(X^k), \quad (4.6)$$

$$\Theta^k = X^{k+1} - X^k, \quad (4.7)$$

$$U^k = B^k \Theta^k \quad (4.8)$$

4.1.7 Levenberg-Marquardt

Another strategy that we considered is the Levenberg-Marquardt method [80, 87] with the search direction determined as:

$$d_k = -(J(X^k)J(X^k)^T + \lambda_k I)^{-1} J(X^k) F(X^k) \quad (4.9)$$

where $\lambda_k > 0$ and the $(J(X^k)J(X^k)^T + \lambda_k I)^{-1}$ is positive definite.

4.1.8 Powell's Dog Leg

In this technique, a strategy is proposed to choose optimization steps based on a parameter called radius of trust region Δ . Three search directions are applied as follows:

$$h_{gn}^k := -J(X^k)^{-1}F(X^k), \quad (4.10)$$

$$h_{sd}^k := -\alpha J(X^k)^T F(X^k), \quad (4.11)$$

$$h_{dl}^k := h_{gn}^k, \quad \text{if } \|h_{gn}^k\| \leq \Delta \quad (4.12)$$

$$h_{dl}^k := \frac{\Delta}{\|h_{sd}^k\|} h_{sd}^k, \quad \text{if } \|\alpha h_{sd}^k\| \geq \Delta \quad (4.13)$$

$$h_{dl}^k := \alpha h_{sd}^k + \beta(h_{gn}^k - \alpha h_{sd}^k), \quad \text{o.w.} \quad (4.14)$$

where

$$\alpha := \frac{\|J(X^k)^T F(X^k)\|^2}{\|J(X^k)J(X^k)^T F(X^k)\|^2} \quad (4.15)$$

Here, we have also used the strategy applied in [95] to choose β and Δ . The general process of mentioned algorithms is illustrated in next sections.

4.1.9 Dijkstra Gaussian Mean Curvature (DGMC) technique

As a heuristic search to find a proper initial value for the normal field, we have devised a technique that consists of two phases. At the first stage, we employ the Dijkstra algorithm [40] to find the nearest point in the outer highlight boundary S to each specular pixel p . This point is denoted as q .

In the second step, we compute the *Gaussian Curvature (GC)* and the *Mean Curvature (MC)* properties of p and q , separately:

$$GC = K_1 \times K_2, \quad MC = \frac{K_1 + K_2}{2} \quad (4.16)$$

where K_1 is the smallest and K_2 is the biggest eigenvalues of the Hessian matrix of the pixel local neighboring area. Having those parameter for both points allows us to define

their local geometric similarity conditions as follows:

$$GC_q = GC_p \pm 5\% \quad (4.17)$$

$$MC_q = MC_p \pm 5\% \quad (4.18)$$

In case that q does not meet similarity conditions, the next nearest member of S to p will be nominated.

4.1.10 Recurrent Optimization Network (RON)

We will introduce an intermittent optimization strategy which allows updating not only the surface normal but also the albedo and roughness parameter. At first, we consider a constant albedo and also a constant roughness so that we can obtain the surface normal. After obtaining the normal field, we update the albedo values. Furthermore, the roughness parameter is updated as well. To this aim, we use the following formulation proposed in [56] to update the roughness parameter m , where we already obtained the depth map z by integrating the normal field.

$$m = \sqrt{\frac{1}{n} \sum_{x=1}^n \sum_{y=1}^n (z(x, y) - \bar{z}(x, y))^2} \quad (4.19)$$

Here, $\bar{z}(x, y)$ is the mean surface depth, and n is the number of pixels. The whole algorithm of mentioned RON procedure is illustrated in Algorithm 2.

Thanks to this concise optimization network with back-tracing steps and DGMC technique which are integrated into the photometric stereo algorithm, our method enables PS to work in diverse applications by removing restrictions such as diffuse objects, orthographic projection, laboratory set up and professional equipment.

Algorithm 2: Recurrent optimization network

output : Normal vectors, colorful albedo, roughness value and the depth map

while *stopping criterion is not satisfied* **do**

1. Compute normal vectors from (4.1) using the proposed DGMC technique and one of Algorithm 3, Algorithm 4 or Algorithm 5 by the constant albedo and roughness
2. Update albedo from (4.1) by the obtained normal field using the same choice of Algorithm 3, Algorithm 4 or Algorithm 5
3. Integrate the normal field using the proposed approach in section 4.1.11 to obtain depth map z
4. Update the roughness value by applying equation (4.19)
5. Update the normal field from (4.1) using the recent albedo and roughness

end

4.1.11 Integration

As the integrator of the normal field, the energy minimization of the following functional should be considered:

$$\mathcal{F}(u) = \iint_{\Omega} \|\nabla z(x, y) - N(x, y)\|^2 dx dy \quad (4.20)$$

which leads to the Euler-Lagrange equation $\Delta z = \nabla \cdot N$.

In order to solve the mentioned minimization problem, we applied Generalized Minimal RESidual (GMRES) with the initial solution of Simchony [119].

4.2 Experiments

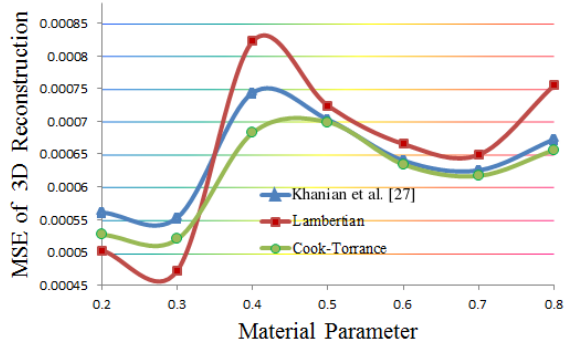
In the case of synthetic experiments, to perform a quantitative evaluation, we will make use of Mean Angular Error of Normal vectors (MAEN) in degrees and also Mean Squared Error (MSE) of depth.

4.2.1 Tests of accuracy on synthetic data

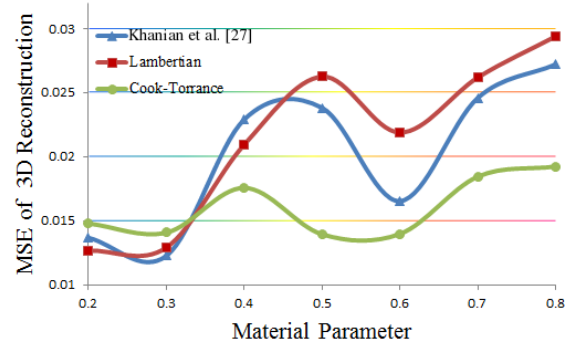
In the following, we discuss our results obtained for the synthetic test images as displayed in Figures 4-6, 4-7 and 4-8. All images are rendered using Blender software. The 3D model of Oldman and Woman images are obtained from [8] and [11] respectively. 3D models of the other synthetic images are publicly available at [12]. All images except *Coping* (256×200) are in the size of 256×256 . All implementations are performed on an Intel Core i7 processor with 8 GB of RAM.

Comparison with other BRDFs:

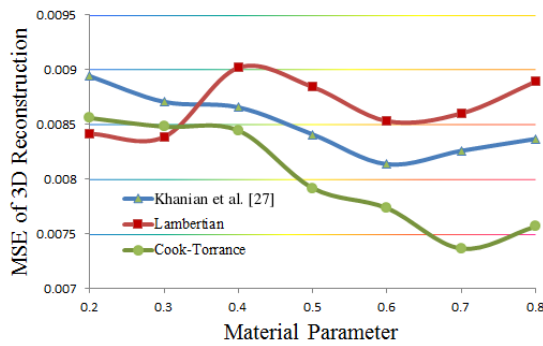
In Fig. 4-3 documenting the first set of experiments, we represent the results of comparison between the recent approach suggested by Khanian et al. [75] and also the results of Lambertian reflectance model (as the most common model applied in PS). The model presented in [75] is based on the Blinn-Phong reflectance. We extended our experiments to a wide range of the specular material parameter $k_s = 1 - k_d$ to evaluate how the varying k_s influence the error of 3D reconstruction. It can be noticed that the output of our method achieves higher rate of accuracy over [75] and Lambertian model as k_s increases. Whereas, there are some fluctuations when $k_d \gg k_s$ or $k_d \ll k_s$. As illustrated, error rates of presented models are almost similar for low values of k_s . However, as k_s increases, the superiority of our model over Lambertian and [75] is significant. This observation agrees well with our motivation for focusing on the high frequency variations and model these components explicitly. One particular interpretation for the sharp ascending trend of the error rate for *Coping* input image (in the case of higher values of k_s parameter) is the presence of the flat background region which leads to high amount of shiny highlighted areas in input images. As for the difference between [75] and Lambertian model, we find that the Lambertian still performs a little better on very low values of k_s ; however, when k_s is larger than 0.3, [75] improves accuracy much more than Lambertian model. As indicated, the MSE comparing our 3D reconstruction and ground truth data, demonstrates that our method is capable of producing accurate 3D reconstructions for the indicated images with specularity. The qualitative evaluations can be also seen as recovered surfaces shown in Fig. 4-6. Our



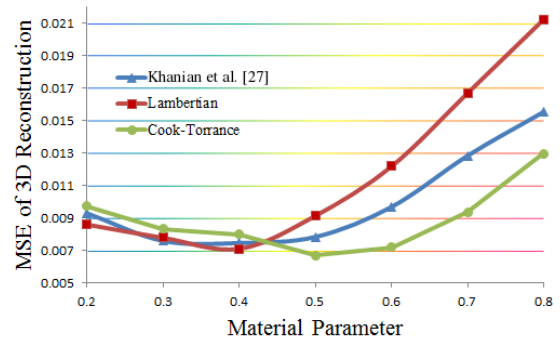
(a) Soldier



(b) Horseman



(c) Fighter



(d) Coping

Figure 4-3: Comparison with [75] and results obtained by Lambertian reflectance (as the most common model applied in PS) in presence of specularity. For a fair evaluation, these experiments are performed against various material parameter (k_s) for different input images [73].

faithful 3D reconstructions with a high amount of details even for tiny surfaces containing low-depth details (e.g. Coping) represent the proficiency of our approach.

Optimization procedures:

As the next experiment shown in Fig. 4-4, we compared various optimization procedures (applied in first and second steps of RON technique) for several surfaces with diverse k_s . It turns out that Dog Leg method achieves higher accuracy in all cases. Although, BFGS and Levenberg approaches follow the same trend for lower values of k_s (Levenberg method can slightly outperform BFGS technique for low values of k_s) BFGS performs better as k_s increases.

Table 4.1: MSE of our 3D shape reconstruction with perspective Cook-Torrance reflectance model in the presence of specularities using different lightening models.

MSE comparison for the applied Lightening models	
PLPS for Oldman	0.007700
DLPS for Oldman	0.012408
PLPS for Woman	0.008439
DLPS for Woman	0.011801
PLPS for Carpenter	0.002580
DLPS for Carpenter	0.003085

Lightening analysis:

To investigate the 3D reconstruction sensitivity w.r.t. the lightening model, we applied different PS frameworks. Quantitative and qualitative results of this investigation are shown in Table 4.1 and Fig. 4-7, respectively. To this end, we evaluated our approach regarding different available lightening conditions. In the first experiment, we consider the distant light source leads to perform directional-light-based PS referred as DLPS. The output images of this simulation resemble the real-world images that are illuminated by sun. The second evaluation is conducted to simulate spatially dependent light resulting in point-wise-light-based PS denoted as PLPS. We observed that the higher accuracy is achieved by applying the point-wise light source and adjusting lightings based on spatially dependent light as presented in (2.30). Fig. 4-7 reveals that DLPS has problem to recover some areas since capturing images by directional-light leads to more shadows as shown in Fig. 4-7. On the other hand, using the point-wise light source in the scene and formulating PLPS can provide the higher amount of accuracy (cf. Table 4.1).

Table 4.2: Our total speed improvement over other methods for all images of Table 4.3. It can be seen that our scheme can decrease consuming time dramatically as well as providing reliable results.

[67]	[21]	[89]	[145]
99.87%	99.90 %	96.31 %	96.63 %

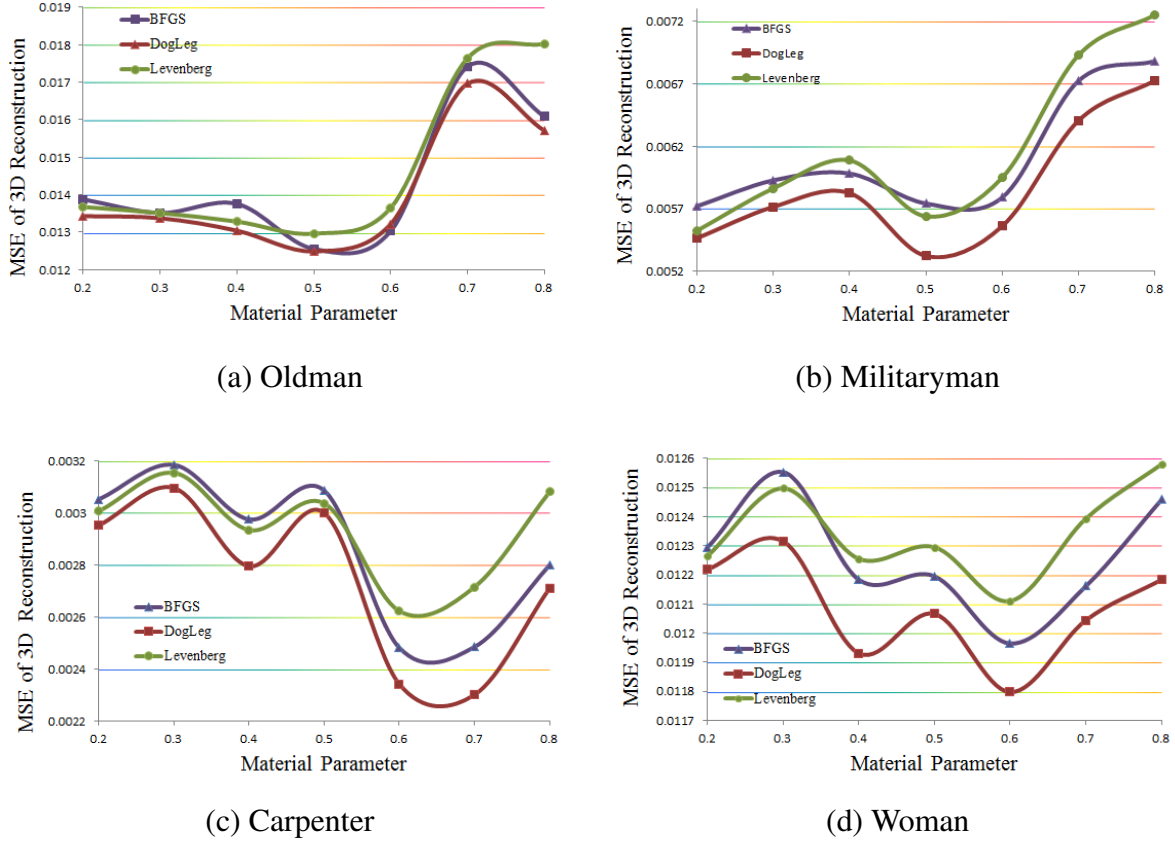


Figure 4-4: MSE of our 3D reconstructions with different optimization procedures for specular surfaces with varying material parameter (k_s) for different input images [73].

Comparison with several PS methods:

As another evaluation shown in Table 4.2 and Table 4.3, we compared our method from different aspects on the set of images with the state-of-the-art approaches: constrained bivariate regression [67], differential ratios [89], dimensionality reduction based on PCA [21] and also least squares-regression [145]. For a fair comparison, normal maps are obtained by the finite-difference of depth map for the methods suggested in [67, 89] (these schemes provide depth, but not normal vectors) and the depth map is obtained by our integrator (section 4.1.11) for [21, 145] (these techniques obtain normal vectors, but not depth map). All methods show worse performances than our approach. One reason is that high-frequency variations are mis-classified and leads to the unpredictable errors. As can be seen in Table 4.3, the proposed strategy dramatically reduces error in computing normal vectors and depth map in comparison with other approaches thanks to our effective techniques of DGMC and RON and capability of handling specularities. The total accuracy of obtained

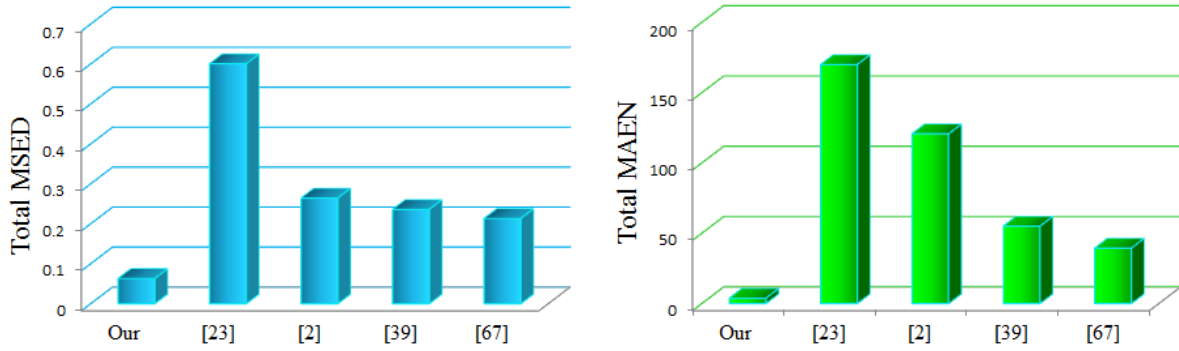


Figure 4-5: Left: Total MSED and Right: Total MAEN gained by our approach and compared methods on all images mentioned in Table 1 [73]. As it can be seen in the graph, the proposed method can be also favored due to its higher accuracy w.r.t both obtained normal vectors and depth map. To illustrate, our technique can provide accuracy improvement in normal field extraction over [67] by 97.78 % and over [21] by 96.87 % . Furthermore, this improvement in accuracy of depth reconstruction by our scheme over [67] is 89.51 % and over [21] is 76.17 %.

normal field and depth map for all images applied in Table 4.3 is shown in Fig. 4-5. For example, we could improve the accuracy in extracted normals by 93.15% over [89] and 90.43 % over [145] and provide higher accuracy in depth reconstruction by 73.29% over [89] and 70.46 % over [145]. Furthermore, the first row in Table 4.3 demonstrates the efficiency of our proposed method. Our method is the first one with significant high speed providing all information of normal vectors, colorful albedo, roughness value and depth map. Our speed improvement over mentioned techniques is represented in Table 4.2. The running time of [89, 145] are comparable, yet their results appear both quantitatively and qualitatively different. The qualitative comparison of our approach with the mentioned methods is shown in Fig. 4-8. The performance ranking of methods is consistent with qualitative results. Our results demonstrate the ability of the proposed scheme to provide a complete reconstruction without any deviation at specularities which outperforms other comparing approaches.

4.2.2 Test of applicability on real-world images

Finally, we evaluated the capability of our approach on a set of real-world images shown in Figures 4-9, 4-10, 4-11, 4-12 as a test of the practical applicability of our method. All the images used in these experiments are captured by the camera of a Samsung Galaxy S5

smartphone and using a flash. For image acquisition the photographed figurine is standing on a table covered by a black tablecloth as shown in Fig. 4-2. The light directions are measured by considering the camera as the origin of the coordinate system. The recovered 3D shape for the first real-world image is presented in Fig. 4-9. In this case, we created a complex scene composed of numerous small statues with different sizes as a challenging set. However, we could reconstruct the surface with correct geometry where relevant details and main features are recovered faithfully. Let us stress that our real-world images are taken without any laboratory equipment or controlled setups. Nevertheless, the reconstruction result is in our opinion very reasonable. This test demonstrates the robustness of our method and its potential for real-world applications even using consumer-level equipment. Moreover, again for this real-world scene, we applied different PS-based lightnings models for a same light source of a flash. Improved quality of the 3D reconstruction with spatially dependent lights (PLPS) shown in Fig. 4-9 (in terms of the height field of depth map and details) can be an advantage offered by this model. More real-world experiments are presented in Figures 4-10 and 4-11. As can be seen, our approach succeeds in producing accurate 3D reconstruction results even for high frequency details of tiny structure of Berlin souvenir statues. The colors of these images are extracted in Fig. 4-12 where the diversity of colors is recovered as well in our non-uniform colorful albedo reconstruction. Let us also point out that our results show the capability of reliable 3D reconstruction for specular materials by using a minimum number of input images (for all synthetic and real-world experiments, we used three input images). Thus we document here that our method works under reasonable practical conditions with inexpensive instruments so that it can be used for many potential applications.

4.3 Summary and conclusion

We presented a real-time robust PS which is also benefiting from perspective Cook-Torrance reflectance model to explicitly handle specularities and remove the limitations of working with diffuse materials or orthographic projection. A RON optimization technique based on DGMC approach is introduced to obtain accurate information of normal vectors, depth,

roughness and colorful albedo. Our proposed RON and DGMC techniques can be applied in any highly non-linear optimization framework as a key point for providing multi-variable solutions and proper initialization. These innovations provide important steps towards a reliable PS. Furthermore, we equipped our model with the spatially dependent Lightening offering more reasonable reconstructions. We have demonstrated the applicability of our method by applying minimum number of input images without any laboratory conditions or facilities. Furthermore, we stretch our approach to its full potential by extracting 3D reconstruction of micro-prints. Our method has the potential to be useful as the basis of future developments.

Table 4.3: Comparison of several PS methods w.r.t Mean Square Error of Depth (MSED), Mean Angular Error of Normal vectors (MAEN in degrees), and Time in seconds. Innovations aggregated in our approach leads to a significant improvement regarding obtained normal vectors and depth as well as computational Time.

Coping				Soldier		
Methods	MAEN	MSED	Time	MAEN	MSED	Time
Our	0.4163	0.0070	0.0485	0.2045	0.0007	0.0320
[67]	22.5992	0.1879	40.9955	29.6263	0.0455	32.1078
[21]	12.9628	0.0424	58.4178	19.8908	0.0205	45.9023
[89]	15.2238	0.0495	1.3645	7.2661	0.0131	1.0775
[145]	11.9886	0.0409	1.4385	5.2925	0.0113	1.1977
Horseman				Fighter		
Methods	MAEN	MSED	Time	MAEN	MSED	Time
Our	0.7241	0.0138	0.0470	0.5948	0.0079	0.0460
[67]	24.1228	0.0511	37.5942	15.4457	0.0279	28.2307
[21]	14.2859	0.0306	52.3987	10.7417	0.0107	39.9016
[89]	10.3619	0.0356	1.2781	4.8770	0.0133	0.9929
[145]	4.2274	0.0424	1.3470	4.8039	0.0114	1.0205
Oldman				Militaryman		
Methods	MAEN	MSED	Time	MAEN	MSED	Time
Our	0.6734	0.0124	0.0466	0.2659	0.0053	0.0429
[67]	18.5054	0.0779	37.3778	24.3848	0.06356	40.4892
[21]	13.2650	0.0400	51.4226	19.9172	0.0430	57.1407
[89]	5.1195	0.0290	1.1693	3.9674	0.0399	1.2868
[145]	3.9751	0.0222	1.3908	5.4562	0.0409	1.4496
woman				Carpenter		
Methods	MAEN	MSED	Time	MAEN	MSED	Time
Our	0.6157	0.0128	0.0393	0.2761	0.00308	0.0420
[67]	12.0515	0.11641	21.0516	23.2172	0.03018	39.9192
[21]	8.5208	0.0538	28.3929	21.1159	0.02331	29.1174
[89]	3.3653	0.04354	1.0492	4.8680	0.01192	0.9669
[145]	2.6305	0.0350	0.7499	1.0557	0.00915	1.4744

4.3.1 Algorithms

Here, the whole algorithms of proposed optimization techniques are presented.

Algorithm 3: Quasi-Newton with BFGS updating

Primary input : X^0 that should be computed using our proposed "DGMC" technique explained in section 5.4. of the paper

Secondary inputs: $\vartheta = 0, k = 0$

output : Normal vectors for photometric stereo

while $\|\nabla\Psi(X^k)\| > \varepsilon$, and $k < k_{max}$ and $\vartheta < \vartheta_{max}$ **do**

Compute $d^k := -(B(X^k)^{-1}) \nabla \Psi(X^k)$

$[X^{k+1}, d\vartheta] \leftarrow$ line search (X^k, d^k) as applied in reference [33] of the paper

$\vartheta \leftarrow \vartheta + d\vartheta$

$\Theta \leftarrow X^{k+1} - X^k$

$Y \leftarrow \nabla\Psi(X^{k+1}) - \nabla\Psi(X^k)$

if $(\Theta)^T Y > \sqrt{\varpi} \|\Theta\|_2 \|Y\|_2$ **then**

$U \leftarrow B(X^k)\Theta$

$B(X^{k+1}) := B(X^k) + \frac{Y^k(Y^k)^T}{(\Theta^k)^T Y^k} - \frac{U^k(U^k)^T}{(\Theta^k)^T U^k}$

end

$k \leftarrow k + 1$

end

Where ϖ is the computer accuracy and $\varepsilon = 1e-12$ is considered.

Algorithm 4: Levenberg-Marquardt

Primary input : X^0 that should be computed using our proposed "DGMC" technique explained in section 5.4. of the paper

Secondary inputs: $\vartheta = 2$, $\lambda_k = \tau \max\{a_{ii}\}$, where $\{a_{ii}\}$ is the set of the diagonal elements of $A = J(X)J(X)^T$, $\Psi(X) = \frac{1}{2}\|F(X)\|_2^2$, $k = 0$

output : Normal vectors for photometric stereo

while $\|J(X^k)^T F(X^k)\|_\infty > \varepsilon_1$, and $k < k_{max}$ **do**

$d^k := -(J(X^k)J(X^k)^T + \lambda_k I)^{-1} J(X^k)^T F(X^k)$

if $\|d^k\| > \varepsilon_2(\|X^k\| + \varepsilon_2)$ **then**

$X^{k+1} \leftarrow X^k + d^k$

$\aleph = \frac{\Psi(X^k) - \Psi(X^{k+1})}{\frac{1}{2}(d^k)^T (\lambda_k(d^k) - J(X^k)^T F(X^k))}$

if $\aleph > 0$ **then**

$\lambda_{k+1} = \lambda_k \max\{\frac{1}{3}, 1 - (2\aleph - 1)^3\}$, $\vartheta = 2$

else

$\lambda_{k+1} = \lambda_k \vartheta$, $\vartheta = 2\vartheta$

end

else

 Exit

end

$k \leftarrow k + 1$

end

Where $\tau = 9e-2$ and $\varepsilon_1 = \varepsilon_2 = 1e-15$ are considered.

Algorithm 5: Powell's Dog Leg

Primary input : X^0 that should be computed using our proposed "DGMC" technique explained in section 5.4. of the paper

Secondary inputs: $\Delta = \Delta_0$, $\Psi(X) = \frac{1}{2}\|F(X)\|_2^2$, $k = 0$

output : Normal vectors for photometric stereo

while $\|F(X^k)\|_\infty > \varepsilon_1$ *and* $\|J(X^k)^T F(X^k)\|_\infty > \varepsilon_2$ *and* $(k < k_{max})$ **do**

$$\alpha = \frac{\|J(X^k)^T F(X^k)\|^2}{\|J(X^k)J(X^k)^T F(X^k)\|^2}$$

$$h_{sd} = -\alpha J(X^k)^T F(X^k)$$

$$h_{gn} = -J(X^k)^{-1} F(X^k)$$

$$h_{dl} = \begin{cases} h_{gn} & \|h_{gn}\| \leq \Delta \\ \frac{\Delta h_{sd}}{\|h_{sd}\|} & \|\alpha h_{sd}\| \geq \Delta \\ \alpha h_{sd} + \beta(h_{gn} - \alpha h_{sd}) & \text{o.w.} \end{cases}$$

if $\|h_{dl}\| > \varepsilon_3(\|X^k\| + \varepsilon_3)$ **then**

$$X^{k+1} \leftarrow X^k + h_{dl}$$

$$\varkappa = \frac{\Psi(X^k) - \Psi(X^{k+1})}{L(0) - L(h_{dl})}$$

if $\varkappa > 0.75$ **then**

$$\Delta = \max \{ \Delta, 3\|h_{dl}\| \}$$

else if $\varkappa < 0.25$ **then**

$$\Delta = \frac{\Delta}{2}$$

end

if $\Delta \leq \varepsilon_3(\|X^k\| + \varepsilon_3)$ **then**

| Exit

end

else

| Exit

end

end

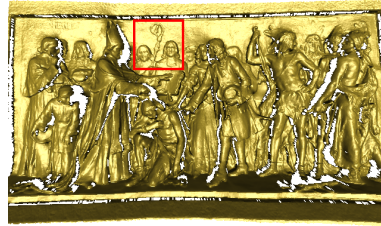
where

$$L(0) - L(h_{dl}) =$$

$$\begin{cases} \Psi(X) & \text{if } h_{dl} = h_{gn} \\ \frac{\Delta \left(2\|\alpha J(X^k)^T F(X^k)\| - \Delta \right)}{2\alpha} & \text{if } h_{dl} = \frac{-\Delta J(X^k)^T F(X^k)}{\|J(X^k)^T F(X^k)\|} \\ \frac{1}{2}\alpha(1 - \beta)^2 \|J(X^k)^T F(X^k)\|^2 + \beta(2 - \beta)F(X) & \text{o.w.} \end{cases}$$



Coping



Horseman



Soldier



Fighter

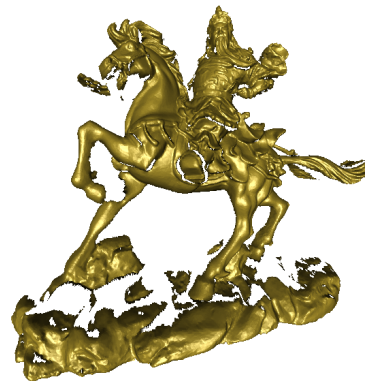


Figure 4-6: Left: input images including specularities [73]. Right: our 3D reconstruction results using perspective Cook-Torrance reflectance model. These results illustrate the capability of the proposed method for providing faithful reconstructions with high frequency details even for the fine details of a tiny surface (e.g. first input image).



Figure 4-7: First row: input specular images with directional light [73]. Second row: our 3D output using DLPS. Third row: input images with pointwise light. Last row: our 3D output with PLPS. Images produced by directional light are imposed by more specularly and shadows in comparison with images with pointwise light. These shadow parts lead to losing some reconstructed area.

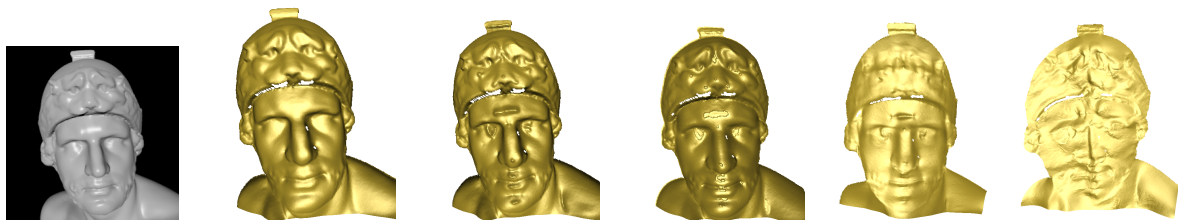


Figure 4-8: Comparison of several PS methods. From left: 1) input image (Militaryman) including high specularities [73], and the 3D results of 2) our reconstruction applying perspective Cook-Torrance reflectance model without any deviation or artifacts, 3) reconstruction of [145], 4) reconstruction of [89] which shows deviations specially around the highly specular areas 5) [21] shows problem in producing complete 3D reconstruction in presence of spcularity and 5) [67] which still includes distortions.

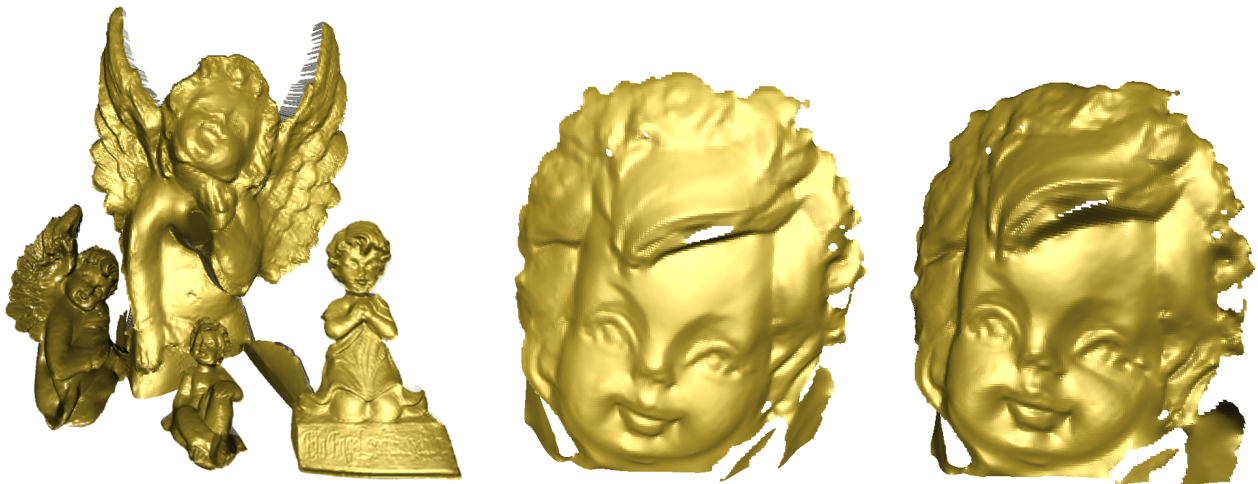


Figure 4-9: From left: 1) input real-world image including several tiny statues captured without laboratory setting cf. Fig. 4-2, 2) our 3D reconstruction with Cook-Torrance reflectance model, 3) our 3D reconstruction using DLPS for the face of the smallest statue in the scene and 4) our 3D reconstruction using PLPS for the same statue [73].

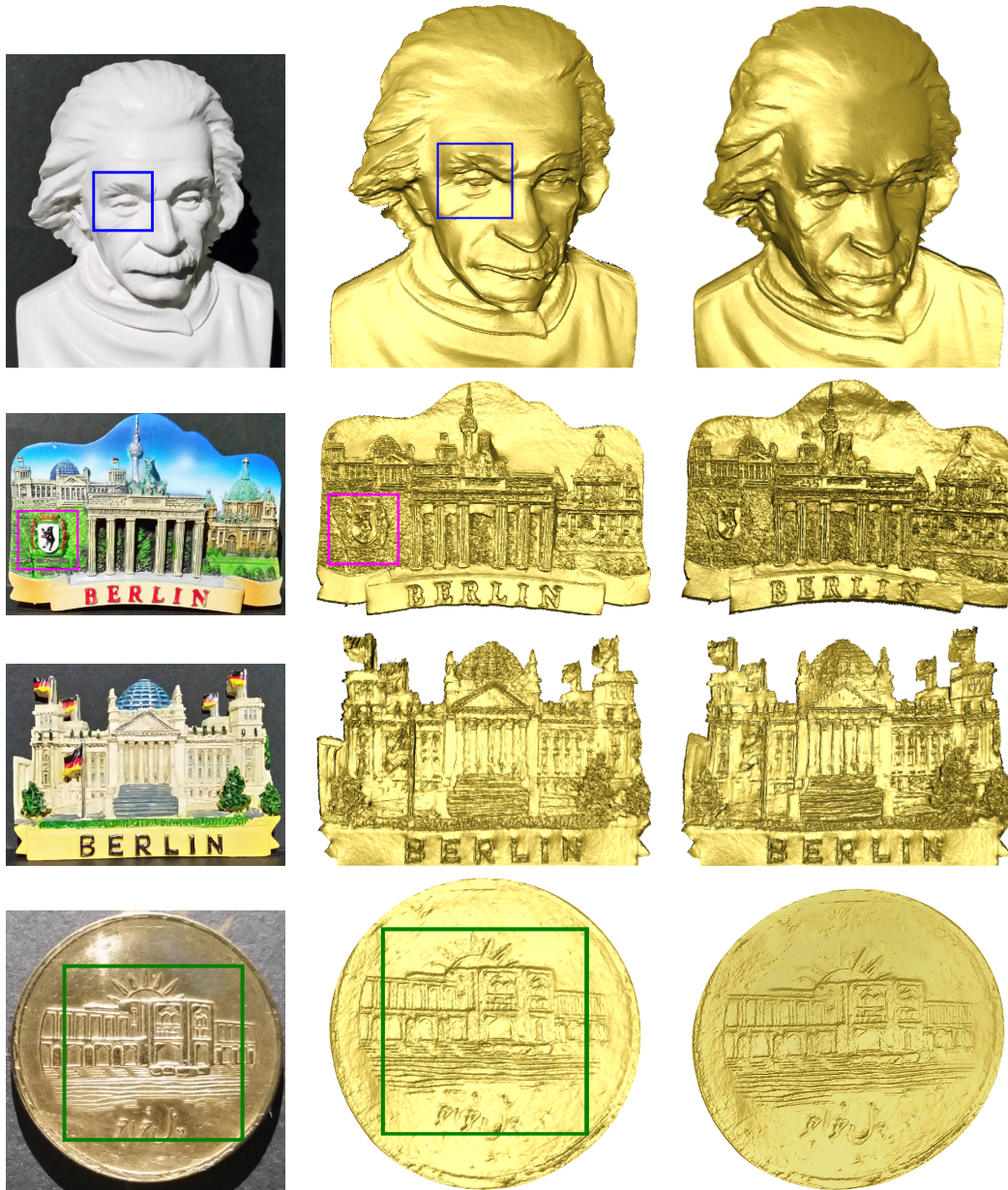


Figure 4-10: More real-world experiments with specularities, first row) input images including Einstein statue, two Berlin souvenir statues and a metallic coin, second row) their 3D reconstruction using DLPS and last row) their 3D reconstruction using PLPS. High quality details and structures recovered by the proposed approach confirm offered advantages [73].

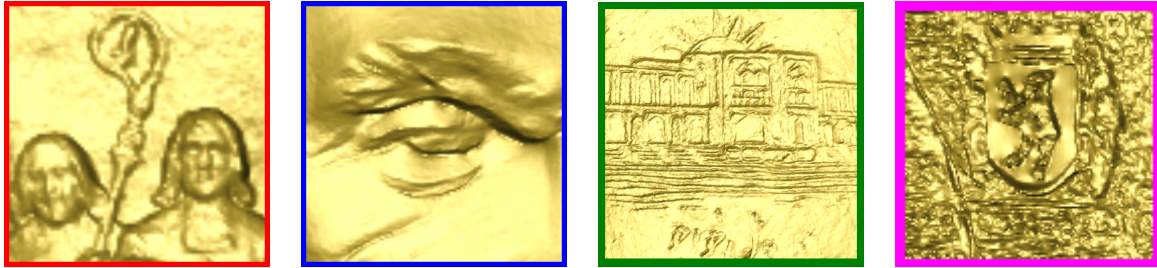


Figure 4-11: Some qualitative evaluations for the proposed method [73]: from left) zooms of our reconstruction results corresponding to the red, blue, green and magenta rectangular areas in Coping as synthetic input and Einstein, Coin and first Berlin statue as the real scenes. It can be seen that our method achieves desirable reconstruction quality even for very fine details.



Figure 4-12: Reconstructed albedo of two color input images shown in Fig. 10. As can be seen, our RON approach can produce non-uniform albedo from images with a variety of colors. Since the albedo is not same across the entire real-world surfaces, our variant albedo extraction is more adjusted for real-world situations. It should be noticed that these statues are very tiny surfaces [73].

Chapter 5

A supervised learning approach for uncalibrated PS using a general reflectance model

5.1 Introduction

5.1.1 Our contributions

In the PS setting, the light source direction should be given in advance. In this section, we extend the task of photometric stereo to the Neural Network (NN) based architecture and the supervised learning framework where the pre-knowledge of the light direction is not necessary. Furthermore, the pixel-wise albedo values are the extra information that will be provided.

5.1.2 Related works

Most works in the area of uncalibrated PS apply Lambertian reflectance equation and are based on factorization technique proposed by Hayakawa [58]. Hayakawa proposed a bilinear modelling based on the Lambertian surface. The presented method needs to consider the constraint like: providing at least 6 pixels with constant or known relative value of the

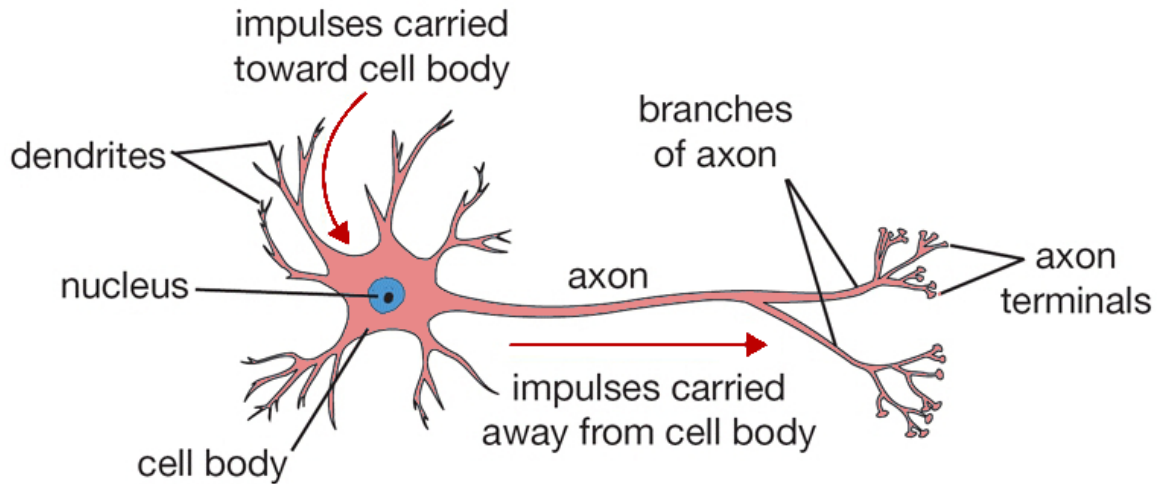


Figure 5-1: Different components of a biological neuron [10]. The figure is redrawn based on [14]

surface reflection. Basri et al. [22] suggested a descriptive photometric model based on the spherical harmonic representation of lighting for Lambertian objects using more than ten images [22]. Other recent works are suggested by [84] using more than one hundred images and [32] which needs pre-knowledge to determine surface normals. The most remarkable work in this area is presented by [83], where a symmetric neural building block is used. In [83], a hybrid reflectance model is proposed that uses two parallel building blocks to simulate diffuse and specular elements of the hybrid model.

In this part, we develop an architecture to utilize a more accurate specular component leading to enhance the quality of the resulting depth and albedo outcomes.

5.1.3 Artificial Neural Networks

With the use of different cortices of the brain, our visual system is able to perform complicated visual recognition tasks. Such an ability is the result of a long natural evolution process in a world where the whole information is in the shape of electromagnetic signals. So, our brain consists of a complex combination of neurons in which each individual neu-

ron can form many of connections with other neurons. Implementation of such an ability in the digital world is a hard task.

Based on the human brain structure, the basic building block of the artificial neural networks is a neuron. As illustrated in Fig. 5-1, a biological neuron is formed by several axon terminals connected via synapses to dendrites of other neurons. During the learning process of the human, the obtained knowledge will be encoded by adjusting connections between neurons. Although, the role of a single neuron in the whole process of the brain should not be neglected. Inter-neuron connections are the place of linear and nonlinear operations that shape the brain elementary computations. If the summation of the incoming signals reaches a predetermined level of value, the neuron will be activated and starts to generate its output. The artificial neuron is designed based on the same concept. It takes several inputs, determines each input impact by its coefficient or *weight*, computes aggregation of weighted inputs and provides the output of its computations if the aggregation output exceeds the predefined threshold of neuron activation. Since the fundamental mathematical aspect of the neural networks is the function estimation, applying a linear activation function such a step function on the linear combination of inputs preserves the linearity, while many interesting functions are nonlinear. Consequently, in modern network architectures, a nonlinear function such as Sigmoid or hyperbolic tangent is used as an activation function

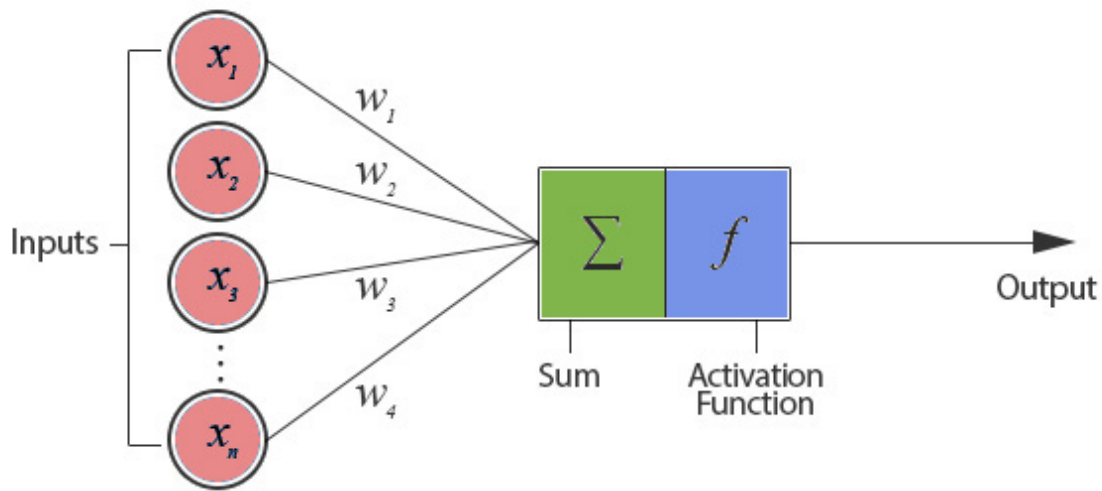


Figure 5-2: The general concept of an artificial neuron [10]. The figure is redrawn based on [15]

to confine the output range of the neuron in a certain interval and provide nonlinear functionality to the neuron. Moreover, there can be an extra input known as *bias* in each neuron that affects the initial level of the desired non-linearity performance. Such a structure of an artificial neuron is shown in Fig. 5-2.

5.1.4 The architecture of our proposed method

The heart of our proposed model is a dual pseudo autoencoder building box where the shape of each sub-network is illustrated in Fig. 5-3. In each module, the first layer is used to normalize the input vector that contains all the pixel values of each input image. The second layer is the albedo applicator, the third and fourth layers are light extractor and normalizer respectively. The last two layers are used to extract and normalize the reflectance map values as the final output of the network that is expected to be the same as the network input.

In the proposed building block, the normalized pixel values at layer one should be divided by pixel-wise albedo values at layer two to generate $\frac{intensity}{albedo}$ values equivalent to $normal \times light$ based on the Lambertian reflectance model. The weight matrix between the second and the third layer (W) should contain the inverse value of the normal vectors to extract light directions in layer three while the weight matrix between the fourth and the fifth layers (N) is expected to include normal values to operate inversely and reproduce $normal \times light$ values as the same value of the reflectance intensities.

5.1.5 Network specular component

While we implement the Lambertian reflectance model in diffuse component of the network, the specular component operates based on the Ashikhmin-Shirley [17] reflectance equation:

$$I(x, y) = \frac{\sqrt{(\eta_u + 1)(\eta_v + 1)}}{8\pi} \frac{\left(\frac{H \cdot N(x, y)}{\|H\| \|N(x, y)\|} \right)^\zeta}{\frac{H \cdot V}{\|H\| \|V\|}} F, \quad (5.1)$$

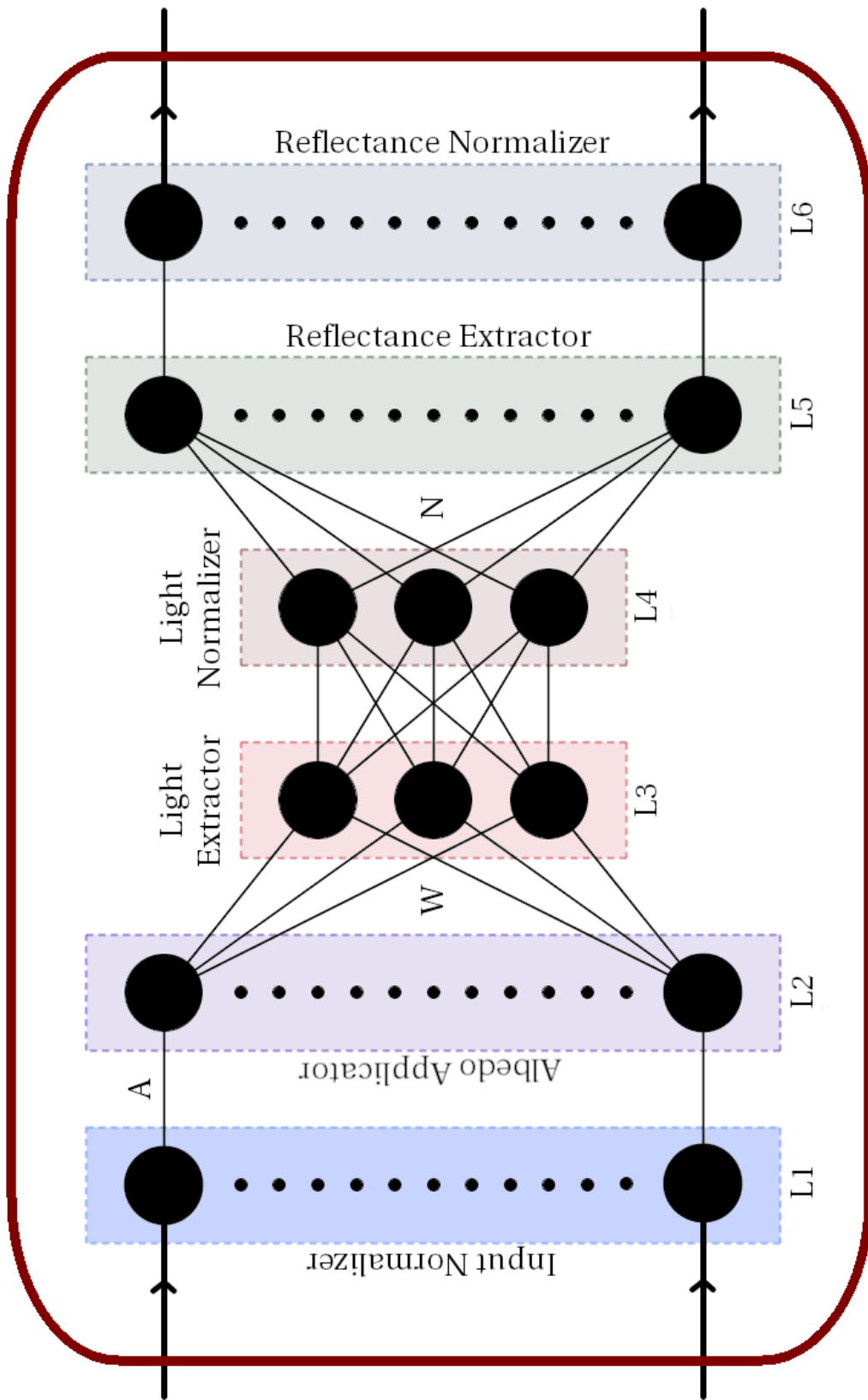


Figure 5-3: Architecture of the proposed symmetric neural building block [73].

$$\zeta = \frac{\eta_u h_u^2 + \eta_v h_v^2}{1 - \left(\frac{H \cdot N(x,y)}{\|H\| \|N(x,y)\|} \right)^2} \quad (5.2)$$

where two exponential factors (η_u, η_v) define the reflectance properties of the surface and in our case have the value of 10 and F is the Fresnel reflection. In this case, the viewing direction V is considered as the orthographic viewing direction whereas the material parameter.

During backpropagation through the specular component, we update surface normal vectors located in N weight matrix by:

$$N_{(t+1)} = N_{(t)} + 2 lr (I - R)^T \Phi, \quad (5.3)$$

$$\Phi = \frac{\sqrt{(\eta_u + 1)(\eta_v + 1)} CF (H^2 N^2 (2 \ln(H \cdot N) - 1) + 1)}{8\pi N (H \cdot N)^d ((1 - (H \cdot N)^2)^2 (H \cdot V))}, \quad (5.4)$$

$$C = \eta_u h_u^2 + \eta_v h_v^2, \quad (5.5)$$

$$d = \frac{C}{((1 - (H \cdot N)^2)^2)}. \quad (5.6)$$

where I and R are the sub-network input and output respectively.

For both diffuse and specular components, in each time step t , the heterogeneous albedo map A is the weight matrix between the first and the second layers and can be obtained as:

$$A_{(t)} = \|N_{(t)}\| \quad (5.7)$$

5.1.6 Network weight initialization

For the network initialization, we exploit the SVD decomposition of the input matrix P containing each input image as a column. The weight matrix N (normal surface values) can be initialized by:

$$P = U \Sigma V^T, \quad N = \frac{S P^T}{A} \quad (5.8)$$

where A is the vector of pixel intensities divided by rough albedo value that can be estimated by surface material information. Finally, the W weight matrix can be initialized by the inverse value of the N weight matrix.

5.1.7 Post-processing step

Since the network will be trained using three input images separately, the outcome of the learning process is an independent set of normal vectors for each input image. Since the network is trained using three input images separately, the outcome of the learning process for each input image is the three components of the normal vectors:

$$img_1 \rightarrow \{N_{x1}, N_{y1}, N_{z1}\},$$

$$img_2 \rightarrow \{N_{x2}, N_{y2}, N_{z2}\},$$

$$img_3 \rightarrow \{N_{x3}, N_{y3}, N_{z3}\}.$$

To generate desired components of the final normal vectors, we used the Principal Component Analysis (PCA) to perform *dimensionality reduction* on the network outputs in following steps:

1) Spatial separation of the surface normal components into *component matrices*:

$$N_x = \{N_{x1}, N_{x2}, N_{x3}\},$$

$$N_y = \{N_{y1}, N_{y2}, N_{y3}\},$$

$$N_z = \{N_{z1}, N_{z2}, N_{z3}\}.$$

2) Calculation of the covariance matrix and its corresponding eigenvectors for the transpose of the above-mentioned component matrices:

$$N_x \rightarrow N_x^T \rightarrow COV_x \rightarrow V_x,$$

$$N_y \rightarrow N_y^T \rightarrow COV_y \rightarrow V_y,$$

$$N_z \rightarrow N_z^T \rightarrow COV_z \rightarrow V_z.$$

As the result, each eigenvector matrix contains the *principal directions* of the equivalent component matrix.

3) Since the final aim is to compute the depth information that is related to the third di-

mension of the data, the third columns of V_x , V_y and V_z should be utilized to project each component matrix on its principal component with respect to the depth direction. The outcomes of this step constitute desired components of the final normal vectors:

$$N1 = N_x \cdot V_x(:, 3),$$

$$N2 = N_y \cdot V_y(:, 3),$$

$$N3 = N_z \cdot V_z(:, 3),$$

$$N_{final} = \{N_1, N_2, N_3\}.$$

5.2 Experiments

To evaluate the strength of the proposed neural framework for depth reconstruction and also pixel-wise albedo approximation in the absence of the light information, we used the model for face reconstruction on Yale Face Database B [50]. Furthermore, we compared our approach with state-of-art techniques. It can be seen in Figure 5-4 that our approach is able to offer faithful 3D reconstruction and also albedo without knowing the light directions. As illustrated in Figure 5-5 the quality of the extracted albedo is clearly increased by our approach and the depth distortion is reduced. It should be noticed that all the results are obtained without any pre-knowledge of the light direction. The network converges very fast after 20 epoch for each input image and the desire learning rate is $5e-3$.

5.3 Summary and conclusion

We presented a neural-network-based framework for uncalibrated PS which is also benefiting from Ashikhmin-Shirley reflectance model. A supervised learning technique is used for the training of our proposed model. These attempts lead to obtain accurate information about depth and also albedo. These innovations provide important improvements towards

uncalibrated PS. Furthermore, we evaluated our model using the comparison with state-of-art approaches and show that our model is able to offer more reasonable reconstructions. The applicability of our method is again shown by applying the minimum number of input images without any information about lights.

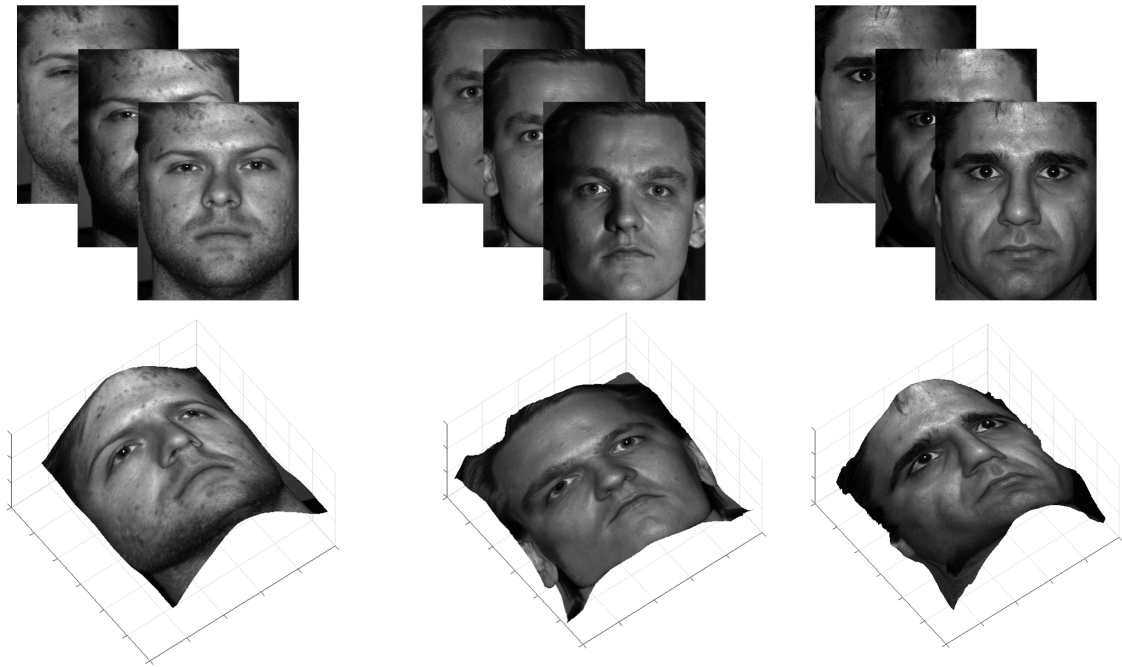


Figure 5-4: 3D and albedo reconstruction using uncalibrated PS with NN-based learning. First row: input images obtained from [50]. Second row: results of our proposed hybrid neural network model for providing the albedo and depth reconstruction .

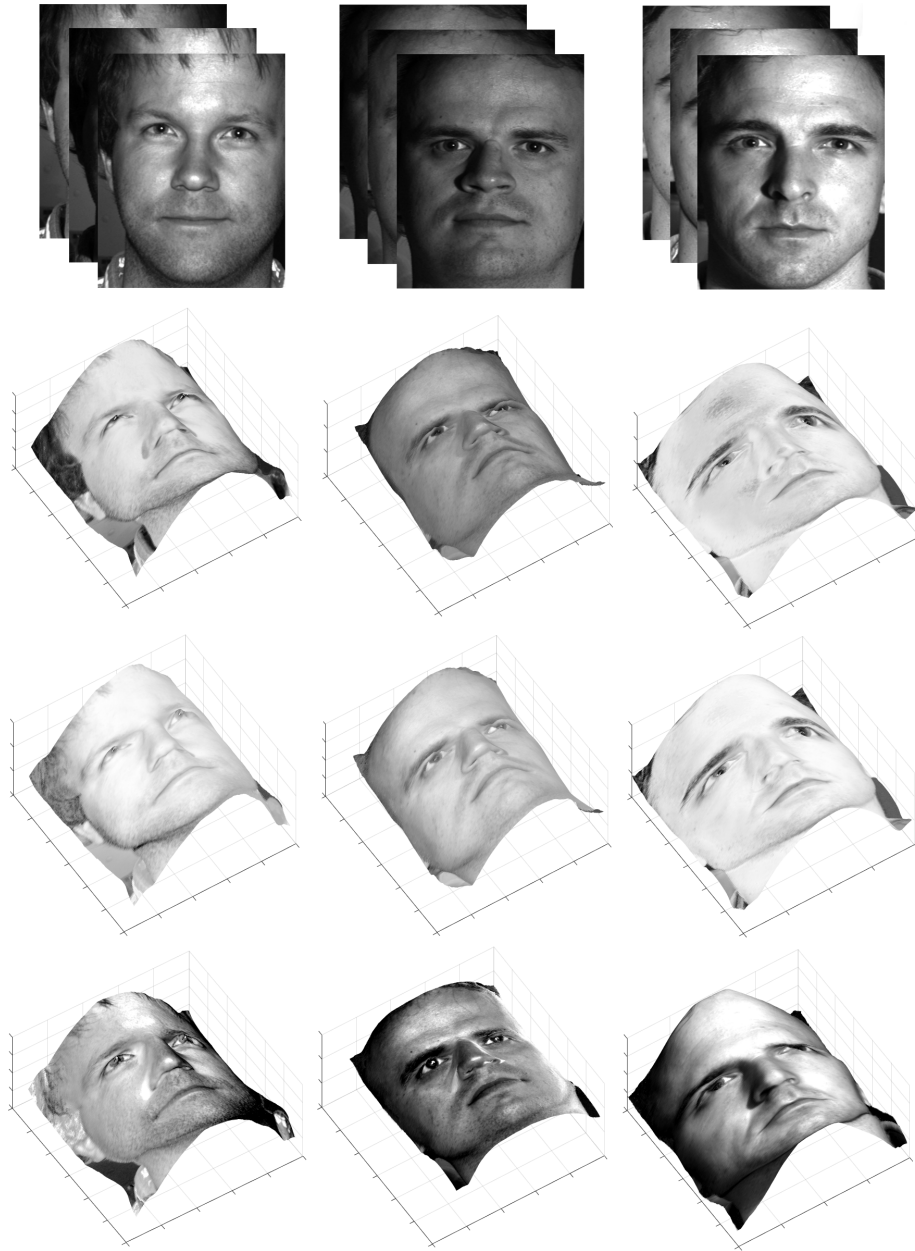


Figure 5-5: Comparison with other neural network based 3D reconstructions. First row: input images [50]. Second row: results of the proposed neural network in [35]. Third row: results of the adaptive neural network suggested in [83]. Last row: results of our proposed hybrid neural network model benefiting from a supervised learning and a general reflectance model which shows more improvements regarding the complex albedo and depth reconstruction provided by our model.

Chapter 6

Deep meta heuristic algorithm for 3D reconstruction

6.1 Introduction

6.1.1 Our contributions

In this chapter, we will devise an innovative model, for the first time, benefiting from a meta-heuristic algorithm and a deep learning framework. Our approach can be applied for finding depth and normal map information of out-door and internet images without any pre-necessary information and calibration.

6.1.2 Artificial intelligence

By the fast progress in different technologies, our new life has drastic changes in comparison with the past. These developments lead to devise effective solutions for real world problems. Nowadays, it is clear that the biological environment acts very productive in solving many tough and complex real world problems [24]. The nature helps scientists to design many equations, formulations and techniques. The science of Artificial Intelligence (AI) is also inspired by the biological world. Artificial intelligence is established in the 1950s with the aim of incorporating the neurology new mathematics in information theory, control theory in cybernetics and digital computers.

Since AI aims at developing and exploring intelligent systems, it focuses on computations. Approaches based on AI are divided into two groups: **Neat** and **Scruffy** techniques. In neat methods, an up-down strategy is applied based on logical process which explains why system works. Although, these neat based procedures had lots of successes, but they show weakness in solving big complex problems regarding some difficulties such as time computation and memory management. In contrast, scruffy algorithms are able to provide solutions for mentioned problems with the reasonable cost.

These schemes apply a descriptive approach to provide a fast solution for complex problems. The most important characteristic of these techniques are using a stochastic process in the decision making and an induction strategy.

6.1.3 Computational intelligence

Computational intelligence is a sub-field in AI which focuses on intelligent and adaptive systems. In fact, this science describes techniques that emphasize strategy-outcome policy. Meta heuristic algorithms are important techniques which are included in the science of computational intelligence as will be discussed in the following.

6.1.4 Meta heuristic algorithms

Meta Heuristic (MH) algorithms belongs to the computational intelligence area which combines natural computing and AI. In 1945, Alan Turing in National Physical Laboratory (NLP) developed an electromechanical machine for automatic computing and in 1948, he provided his innovative theories in learning, machine intelligence, neural networks and evolutionary computation in his reports at NLP. The 1960s and 1970s are two important decades in developing evolutionary algorithms. MH techniques are the results of recent attempts and bio- searches for finding the solution of many complex problems in real world, where classic methods apply just the information existing in the problem and they are not able to overcome tough problems. The robustness of MH methods to the dynamic changes in the environment leads to the variety of their applications such as: Operation of power

systems and control [34, 140, 20], NP-Hard combinatorial problems [47, 27], schedule problems [146], chemical process [31]. They are applicable for both continuous and discontinuous problems. In following, a brief history of important MH algorithms will be provided.

Genetic algorithms

At first, Holland [62] developed Genetic Algorithms in university of Michigan. He presented complex adaptive systems (CAS) studies in 1962. He was the first one who used recombination and mutation techniques for modelling CAS systems. In general, GA is based on the following operations: selection operation for choosing the best parents between the current generations, reproduction operation for producing a new child and mutation operation for changing the gene.

Simple hill-climbing

Rechenberg [107] proposed the evolution technology in technical university of Berlin which was applied in aerospace engineering. Then Bienert joined him and made an automatic experimental device which worked with the mutation and selection rules. This method was a simple hill-climbing with randomization.

Simulated annealing

In 1980s and 1990s , the remarkable changes emerged in meta heuristic algorithms. Simulated Annealing is proposed by Kirkpatrick et al. [76] inspiring from gradual cold processing of metals in which the substance is heated above its melting temperature and then gradually cools to produce the crystalline lattice with the minimum level of energy and entropy.

Ant colony

In 1992, Dorigo devised an Ant Colony optimization in his thesis inspired by the pheromone trail laying of ant colonies. His work is later published in [41]. It is a population based

search technique. This approach is a good example of a cooperation in which the agent which are not inherently capable elements can get results through a collaboration.

Particle swarm

Particle optimization is also another technique provided by Kennedy et al. [72] in 1999. This stochastic algorithm is inspired by social behavior of bird flocking or fish schooling [44]. In fact, it uses two separate procedures: a. **Swarm intelligence**: based on the interactions, habits and repetitive behavior of birds and fishes. b. **Evolutionary computation**: based on the movement of particles. Each particle is moving with a certain speed.

Tabu search

Tabu search is the first technique based on the memory structures proposed by Glover [52]. This approach consists of three important elements: **search space, neighborhood structures and search memory**.

Harmony search

In 2001, Gemm et al. [49] developed harmony search scheme as a special case of evolution strategy and is inspired by the improvisation process of jazz musicians. Harmony search has variety of applications such as water distribution, transport systems, geotechnical analysis, biomedical engineering and prediction of energy demands.

Artificial bee colony

Artificial bee colony algorithm is proposed by Karaboga [71] in 2005. This scheme works based on stochastic movement, information transmission and ranking performance of bees for searching food. Some of its applications are: training of artificial neural networks, time modelling of manufacturing machines, information clustering and adjusting fuzzy logic in robotic science.

Cuckoo search

In 2009, Yang et al. [150] devised a strategy inspired by the obligate brood parasitism of cuckoos which lay their eggs in the nests of other host birds. As proved in [150], this algorithm is more efficient than most of other MA techniques. In our approach, we will develop a 3D reconstruction based on the Cuckoo search and a deep learning architecture.

6.2 Cuckoo optimization algorithm for 3D reconstruction

Cuckoo algorithm is inspired by the cuckoo life. The cuckoo types include the common or European cuckoo, roadrunners, koels, malkohas, couas, coucals and anis. Coucals and anis are usually live as distinct families composed of Centropodidae and Crotophagidae.

Cuckoos are generally medium-sized slender birds. Most of them live in trees, but the habitat of a minority is the ground. They extended as an international distribution, although a majority of species are tropical kind.

Some species are migratory. Cuckoos feed on insects, insect larvae, other animals and fruits. Some species are brood parasites, laying their eggs in the nests of other species. Cuckoos have been existed in human culture for thousands of years (in Greek mythology). In Europe, cuckoos appear in spring. In India, cuckoos are sacred to Kamadeva (the god of desire and longing), while in Japan, the cuckoo symbolises unrequited love.

6.2.1 Brood parasitism for reproduction

Birds have the same approach for reproduction. They lay eggs in a protective shell and incubate them. Some eggs are big and make problems for the female with carrying and caring. In addition, finding a safe place and hatching are hard duties. In general, motherhood is not an easy task. Consequently, some birds found a way which is not fair. Their way is named brood parasites. They never build their own nests and lay their eggs in the nest of other bird. One of the experts in this area is cuckoo with a policy of tough cruel deception. They have different strategies for putting their egg into a host nest, but they are common in these points: stealth, surprise and speed. In addition, they choose their strategy based on

the host defensive strategies. Female cuckoos remove one egg laid by the host and lays her egg instead. This process takes just some seconds (about 10 seconds), but in some cases, males deceive hosts and make them away from their nests so that the female can put her egg in the host nest.

Cuckoo can mimic the color and pattern of host eggs for her own eggs [FIG]. Some other species of cuckoo lay eggs which are dark in color when their hosts' eggs are light [39]. This is a trick to hide the egg from the host, and is exhibited in cuckoos that parasitize hosts with dark, domed nests as a trick to hide the egg from the host. Some European cuckoos that females will lay their egg in the nest of a host that has eggs looking similar to its own [18]. Some birds learn to recognize the cuckoo eggs hidden in their nest and throw it out or even leave their nest because cuckoos completely destroy the host's clutch if they reject the cuckoo egg [39].

As a result, cuckoos constantly try to improve their mimicry policy (cf. 6-1), in return hosts try to be expert in recognizing cuckoo eggs. This attempt is going on as a race to survive [6].

Most species of cuckoo are sedentary, but some of them immigrate regularly to warmer climates in the winter and others immigrate partially around their range. After putting the egg in the host nest, the cuckoo chick hatches after 11–13 days and evicts all host chickens from host nests. The chick rolls the host eggs out of the nest by pushing them using his (her) back. If host eggs born before the cuckoo's, the cuckoo chick again pushes host chicks out of the nest. Then the cuckoo chick forces the host to provide food by constantly screaming.

6.2.2 Cuckoo algorithm

Another point about cuckoo life is that they try constantly to find better position. It means that the new generation of cuckoos tries to follow the successful cuckoos and move to the regions with the lowest possibility of detecting their eggs. So they converge to locations in where more cuckoo eggs survived. As a result, some principles set in this algorithm:

- Each cuckoo lays one egg in the host nest and choose the nest randomly.



Figure 6-1: Cuckoo mimics host birds eggs [3]. Two top images are examples of a successful mimicking and the down image still needs more attempts to provide more similarities.

- The best nest with the highest survived eggs will transfer to the next generation.
- The probability of detecting egg by the host is $p_c \in [0, 1]$.

The goal is finding a new habitat. A habitat is an array $X \in \mathbb{R}^n$. The value of each habitat is shown by $F(X) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and we aim at maximizing the value function. To produce a candidate initial habitat, some random numbers can be provided. For finding the next habitat, cuckoos choose a **Lévy flight** defined as:

$$X^{k+1} := (X^k) \oplus \beta \text{Lévy}(\chi) \quad (6.1)$$

with

$$\text{Lévy}(\chi) = R\psi\chi \quad (6.2)$$

where χ is defined as:

$$\chi = (X^k - X_{best}^k) \quad (6.3)$$

and

$$R \sim N(0, \sigma^2) \quad (6.4)$$

where ψ is defined as:

$$\psi = \frac{u}{|\nu|^{\frac{1}{\beta}}} \quad (6.5)$$

where

$$u \sim N(0, \sigma_u^2) \quad (6.6)$$

and

$$\sigma_u = \left[\frac{\Gamma(1 + \beta) \sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \beta(2)^{\frac{\beta-1}{2}}} \right]^{\frac{1}{\beta}} \quad (6.7)$$

and for the ν :

$$\nu \sim N(0, \sigma_\nu^2) \quad (6.8)$$

with

$$\sigma_\nu = 1 \quad (6.9)$$

\oplus indicates the entry wise multiplication. The step lengths defined in the Lévy have a probability distribution that is heavy tailed. We constitute our system of equations $F(X)$ using Blinn-Phong model. The whole algorithm of cuckoo is illustrated in Algorithm 6. For more discussion about Lévy flight, [82] can be suggested. By this way, new solutions will be generated by Lévy fight around the best solutions achieved so far.

6.2.3 Convolutional Neural Networks (CNNs)

In ordinary feed-forward networks, each node of a hidden layer is fully connected to all the neurons of the previous layer. Apparently, such a connection strategy is computationally expensive when dealing with large color images. For example, when we convert an ordinary color image with the common size of 512×512 into a one-dimensional input vector, this leads to $512 \times 512 \times 3 = 786,432$ input neurons. In deep architectures, we would

Algorithm 6: Cuckoo search algorithm

```
1 while stopping criterion is not satisfied do
    1. Consider the objective function  $F(X) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 
    2. Generate the initial population with  $n$  habitats  $X \in \mathbb{R}^n$ 
    3. Move the initial population using Lévy flight
    4. Choose a habitat  $j$  randomly
    5. Compute the value of habitat  $j$  and  $i$  ( $F(X_i), F(X_j)$ )
    6. If the value of habitat  $j$  is better than  $i$ , accept  $j$  as the new solution
    7. A fraction ( $p_c$ ) of worse habitats can be abandoned and keep the best habitats
    8. Rank the habitats and choose the best one
2 end
```

almost certainly want to have lots of neurons over multiple layers, so the number of the network connections would add up quite quickly.

In contrast, a Convolutional Neural Network (CNN) architecture [77, 120, 127, 59] includes convolutional, pooling and non-linearity layers optionally followed by a few fully connected layers. The network input is normally a 3D matrix with $H \times W \times D$ dimensionality where H stands for Height, W stands for width and D stands for the depth of the input matrix.

As illustrated in Fig. 6-2, in a convolution/pooling block, each neuron has selective connections to a small, local region of the preceding layer and weights of these connections form a filter (kernel) that will be changed during the training phase of the network. At the time of the convolution process, each filter traverses the entire scope of the input matrix to generate its own *feature map*. In practice, filters often have the square shape of size 1×1 , 3×3 or 5×5 . Three hyperparameters control the size of the output volume: *depth*, *stride*, and *zero-padding*.

Depth corresponds to the number of filters in a convolutional layer, where each filter learns to look for something different in the input. For example, if the first convolutional layer takes as input the raw image, then different neurons along the depth dimension may activate

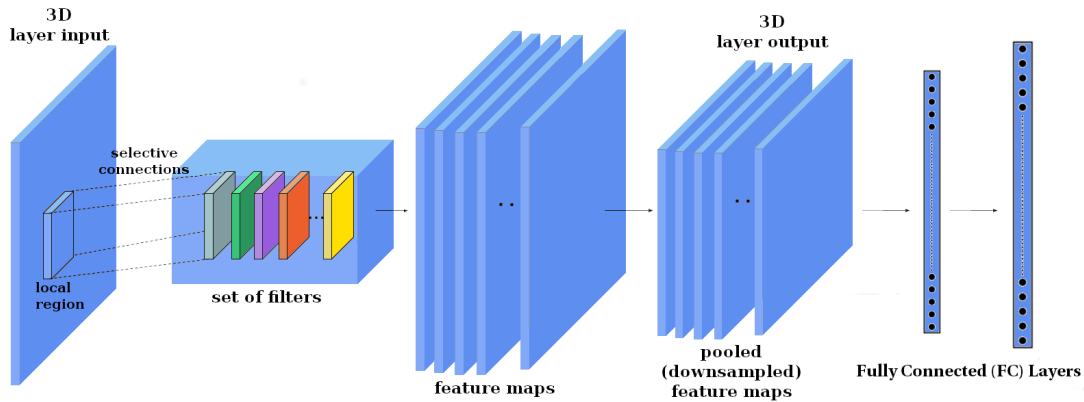


Figure 6-2: A repaint [73] of the general architecture of a Convolutional Neural Network (CNN).

in presence of various oriented edges, or blobs of color. The distance between consecutive applications of a filter over the input scope is called *stride*. When the stride is n , we move the filters n pixel at a time. The bigger the stride, the smaller the output size. Sometimes it will be convenient to pad the input volume with zeros around the border. The size of this *zero-padding* operation enables us to control the spatial size of the output volumes. It also makes it possible to preserve the spatial size of the input volume, so, the input and output height and width remain unchanged. An important property of a convolution layer is that all spatial locations share the same convolution kernel, which greatly reduces the number of parameters of the layer. Because of sharing parameters, a kernel that is specialized to detect a pattern will detect it anywhere in the image.

As mentioned before, it is a convention to apply a non-linear function to grant non-linearity to a system that basically performs linear operations. In comparison with Sigmoid and hyperbolic tangent activation functions, *Rectified Linear Unit* (ReLU) is able to greatly accelerate the convergence of stochastic gradient descent due to its linear, non-saturating form. Moreover, it can be implemented by simply thresholding the matrix of activations at zero. It also helps to alleviate the vanishing gradient problem, which is the issue where the lower layers of the network train very slowly because the gradient decreases exponentially through the layers. The ReLU layer applies the function $f(x) = \max(0, x)$ to all of the values in the input volume. In basic terms, this layer blocks all the negative activations.

Finally, to constantly reduce the dimension of feature maps, sharpen the located features,

and control the amount of the computation in the network, it is common to perform a *pooling* operation after a convolutional layer. Typically, a pooling layer uses the maximum or the average operation to resize the dimension of the input and does not include any parameter. In *max pooling*, the pooling operator maps a sub-region to its maximum value, while the *average pooling* maps a sub-region to its average value. Furthermore, pooling helps to make the representation approximately invariant to small translation, rotation and shifting. In this way, CNNs are able to encode a wide variety of visual features via their specific structure during the learning process. This kind of neural networks is usually used for visual classification, similarity detection (clustering), and object recognition

6.2.4 Deep learning architecture

During last couple of years, CNNs have shown incredible promise in a wide range of well-known computer vision tasks such as object detection, image classification and region segmentation. More recently, their application is extended for learning visual interpretations including depth and surface normals [108, 123].

Now we provide a proper initialization for the cuckoo algorithm and develop it to the PS such that it can overcome complexities like lack of information about light and 3D reconstruction from out-door or internet images. We employed the proposed multi-scale deep neural network of [43] (cf. Fig. 6-3) that is dedicated to estimate surface normals and also depth by regressing from one input image to depth and normal maps in real-time. At the first module, a cascade of convolutional and fully connected layers provides a coarse global prediction as a set of spatially-varying features that are extracted from the whole image area. The task of the second scale that consists of five convolutional layers is to combine a more detailed but narrower view of the image with the full-image information supplied by the previous module. The third and the final scale includes four convolutional layers and predicts x , y and Z components of the normal at each pixel. To this aim, it employs an element-wise objective function to compare the predicted normal of the pixel with the ground truth:

$$L(N, N^*) = -\frac{1}{m * n} \sum_i N_i \cdot N_i^* = -\frac{1}{n} N \cdot N^* \quad (6.10)$$

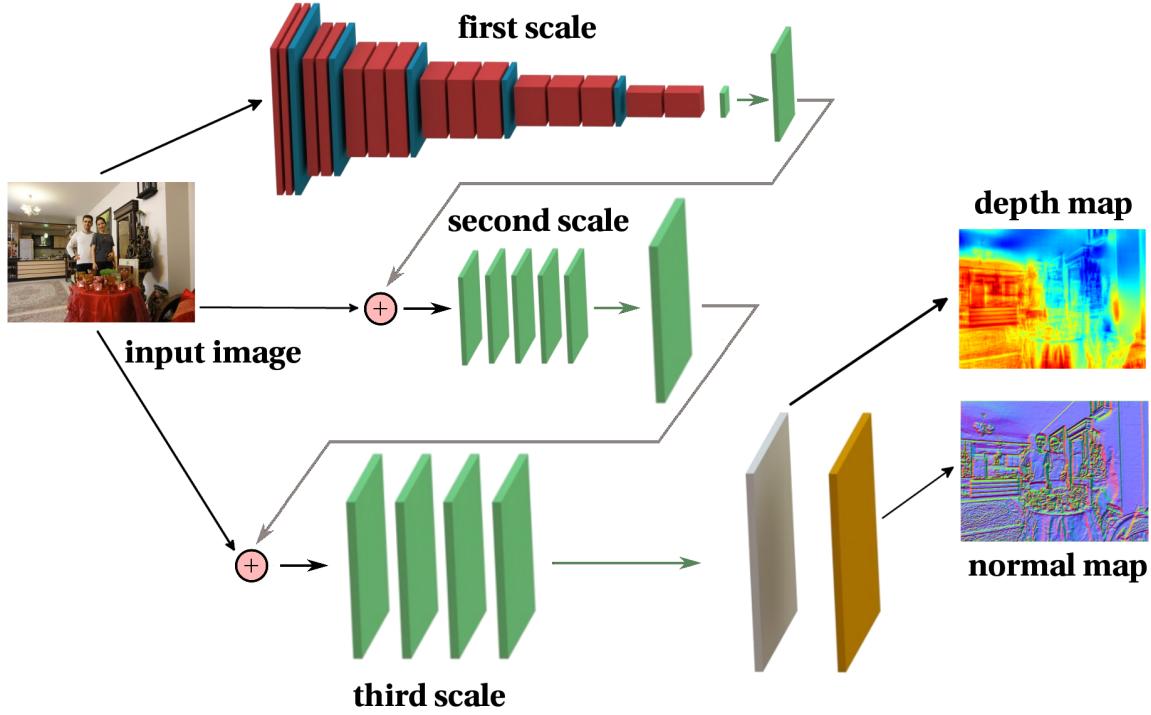


Figure 6-3: Architecture of the deep model that we used to provide initial values of the depth and normal estimation for the cuckoo algorithm [73].

where N and N^* are predicted and ground truth normal vectors, i indicates the valid pixels for normal estimation and $m * n$ is the number of valid pixels.

For the depth reconstruction, the following formulation will be applied:

$$L(Z, Z^*) = \frac{1}{m * n} \sum_i D_i^2 - \frac{1}{2(m * n)^2} \left(\sum_i D_i \right)^2 + \frac{1}{m * n} \sum_i (\nabla_x D_i^2 + \nabla_y D_i^2) \quad (6.11)$$

where, $D = Z - Z^*$ as the obtained and ground truth depth maps. $m * n$ is the number of pixels. $\nabla_x D_i^2$ and $\nabla_y D_i^2$ are the horizontal and vertical gradients.

6.2.5 Experiments

In this section, we will represent our results from both the proposed deep architecture and meta heuristic PS. We will demonstrate our algorithms on real-world scenes of internet images, out-door images and even the image captured from the Moon surface. It should be mentioned that our deep model uses just one input image to produce normal and depth

maps simultaneously, while for the depth reconstruction by PS we should use three input images. The step size β is considered 0.01, p_c is 0.4 for the cuckoo algorithm. Results of the proposed deep architecture are shown in Figs. 6-4, 6-5, 6-6, 6-7 regarding the output normals, depth map and 3D reconstructions. In addition, the outputs of our meta heuristic PS are represented in Figs. 6-8, 6-9, 6-10. Although our deep network succeeds in producing outstanding results using just one input internet image without any information about light direction, scene or calibration, but we can improve its results using our meta heuristic PS specially for tiny surfaces as it is shown in Fig. 6-11.

6.2.6 Summary and conclusion

We presented, for the first time, a technique for the 3D reconstruction from out-door and internet images, where we do not have any information about the light direction, depth values, scene, etc. To this aim, we devised a deep network providing normals and depth values using an input image, then we applied a meta heuristic algorithm on the network output using PS and could improve the results of our deep network. We demonstrated the applicability of our approach using different real-world internet images and also on the moon image. For the future work, we aim at devising a deep network for applying directly PS.

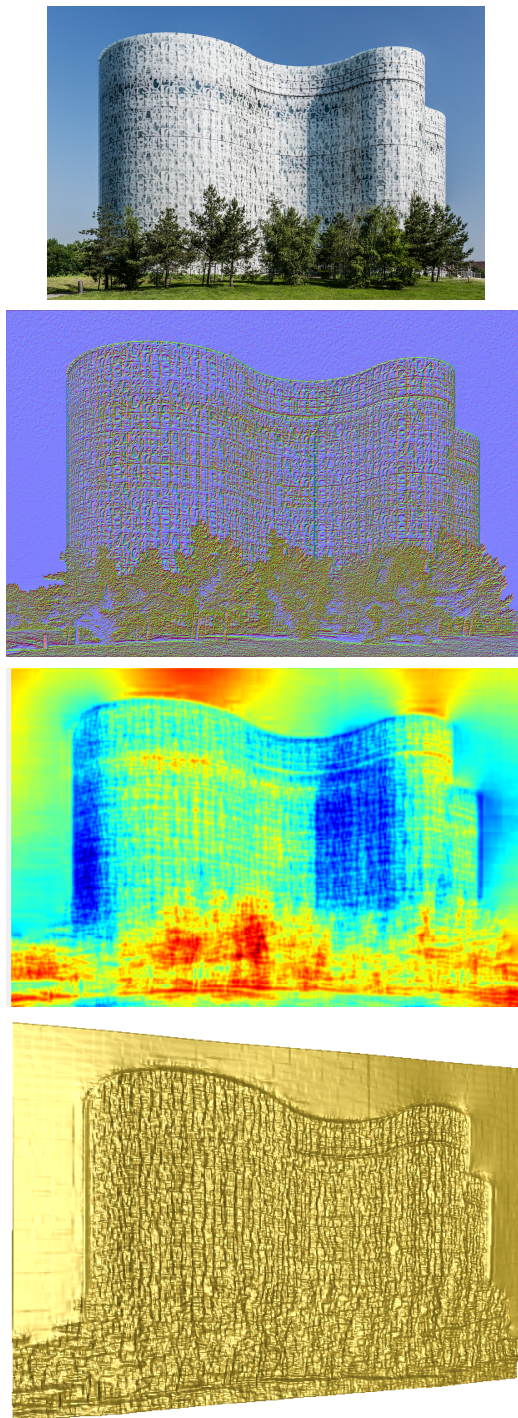


Figure 6-4: First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth maps and 3D reconstruction obtained by the proposed deep architecture model. These results gained by one input image without any information about light direction or scene. As can be seen our approach can recover structures successfully which agrees well with the input image.

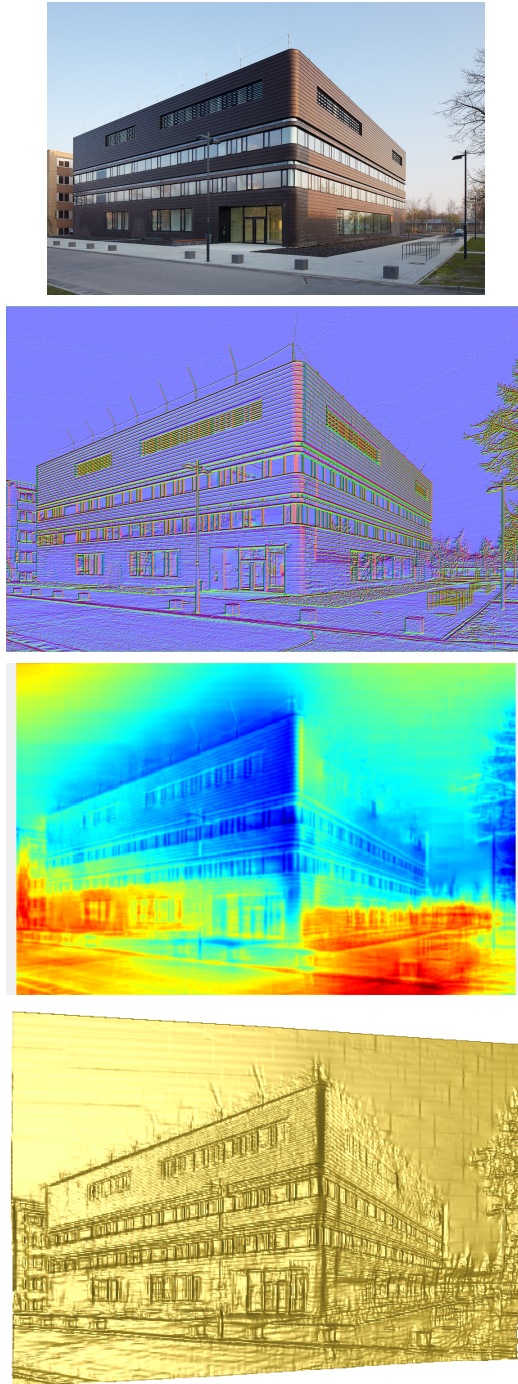


Figure 6-5: First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth map and 3D reconstruction obtained by the proposed deep architecture model.

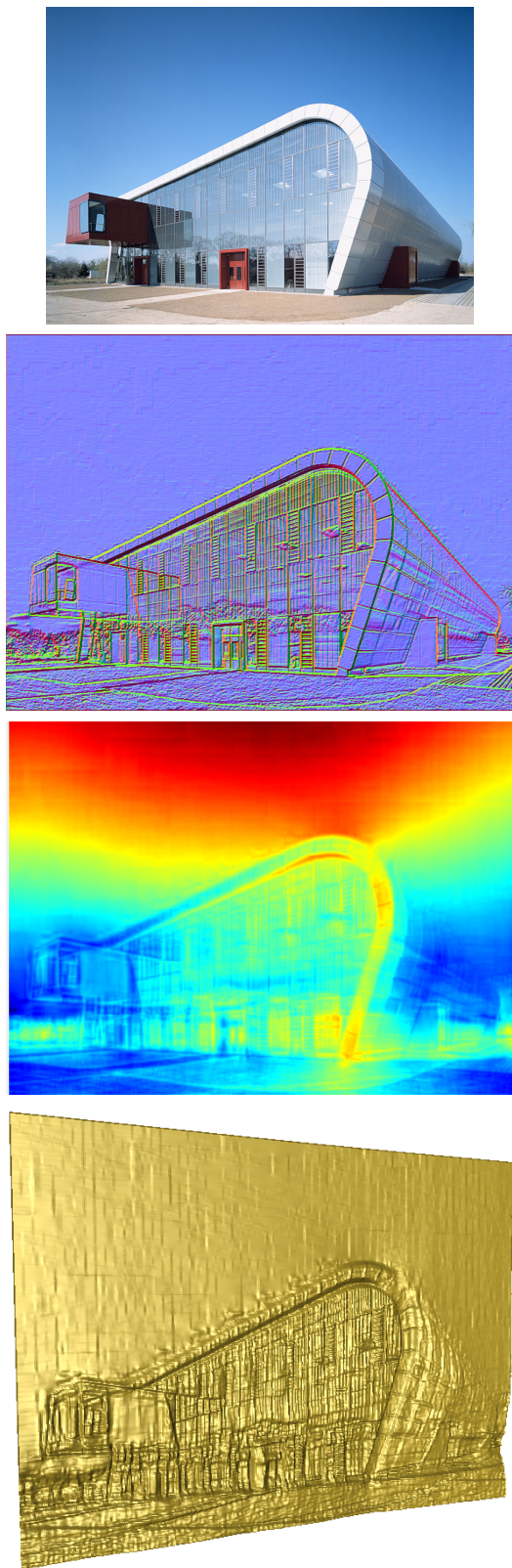


Figure 6-6: First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth map and 3D reconstruction obtained by the proposed deep architecture model.

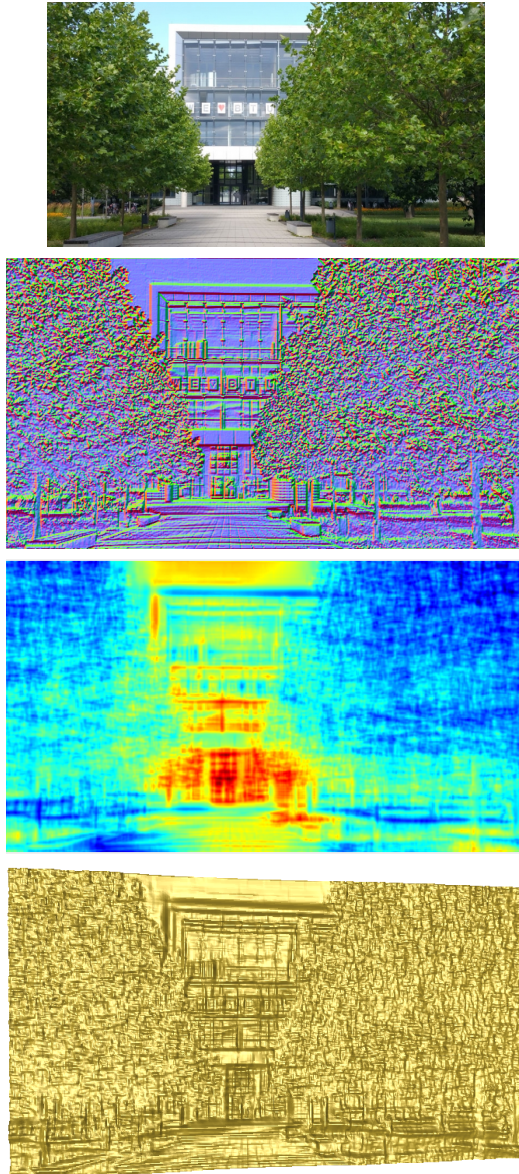


Figure 6-7: First row: one input internet real-world image [73]. Second, third and last rows: our normals, depth map and 3D reconstruction obtained by the proposed deep architecture model.

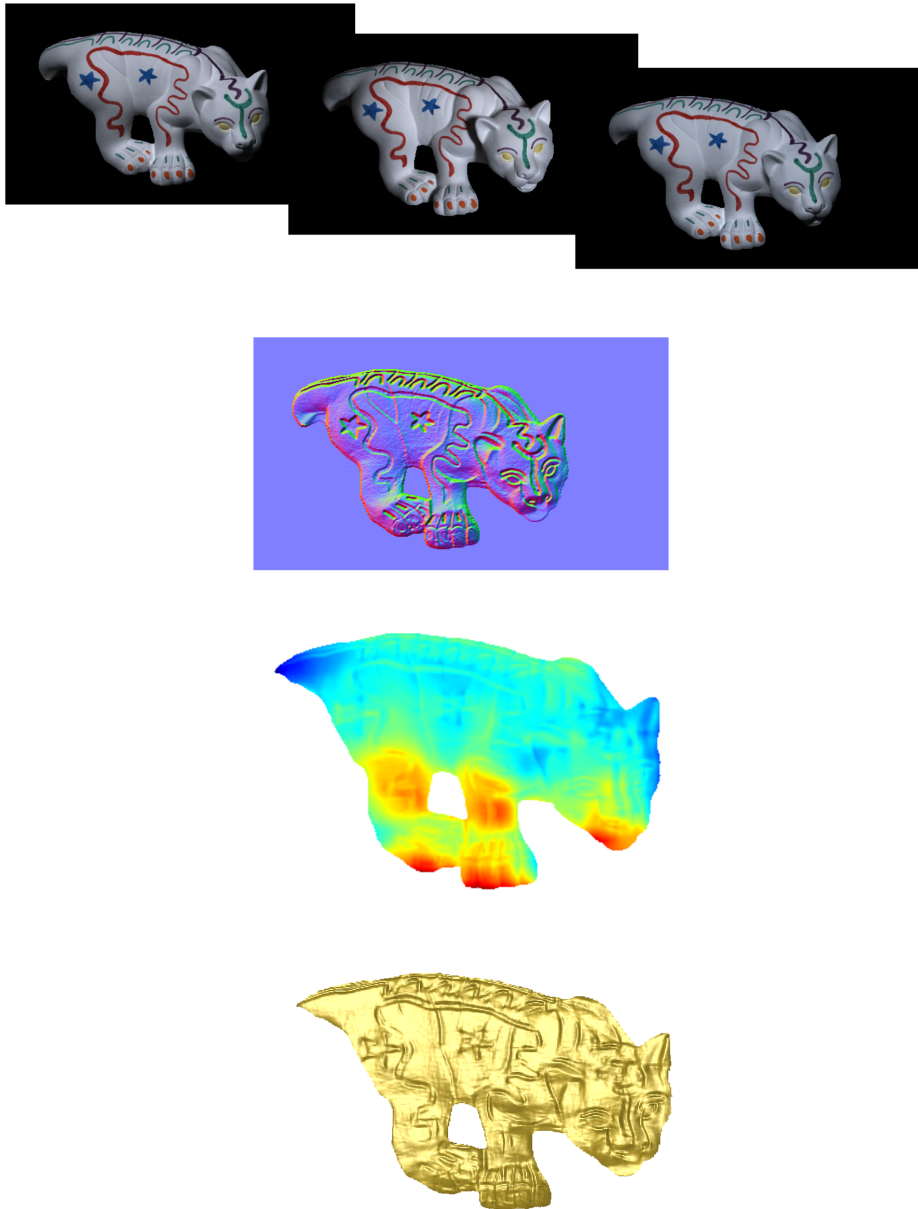


Figure 6-8: First row: three input real-world images from MIT Intrinsic Images [55]. Second, third and last rows: our normal, depth map and 3D reconstruction produced by the proposed PS model. These results obtained without any information about the light direction, scene, etc.. As can be seen our proposed metaheuristic deep architecture can provide high qualified results for uncalibrated PS.

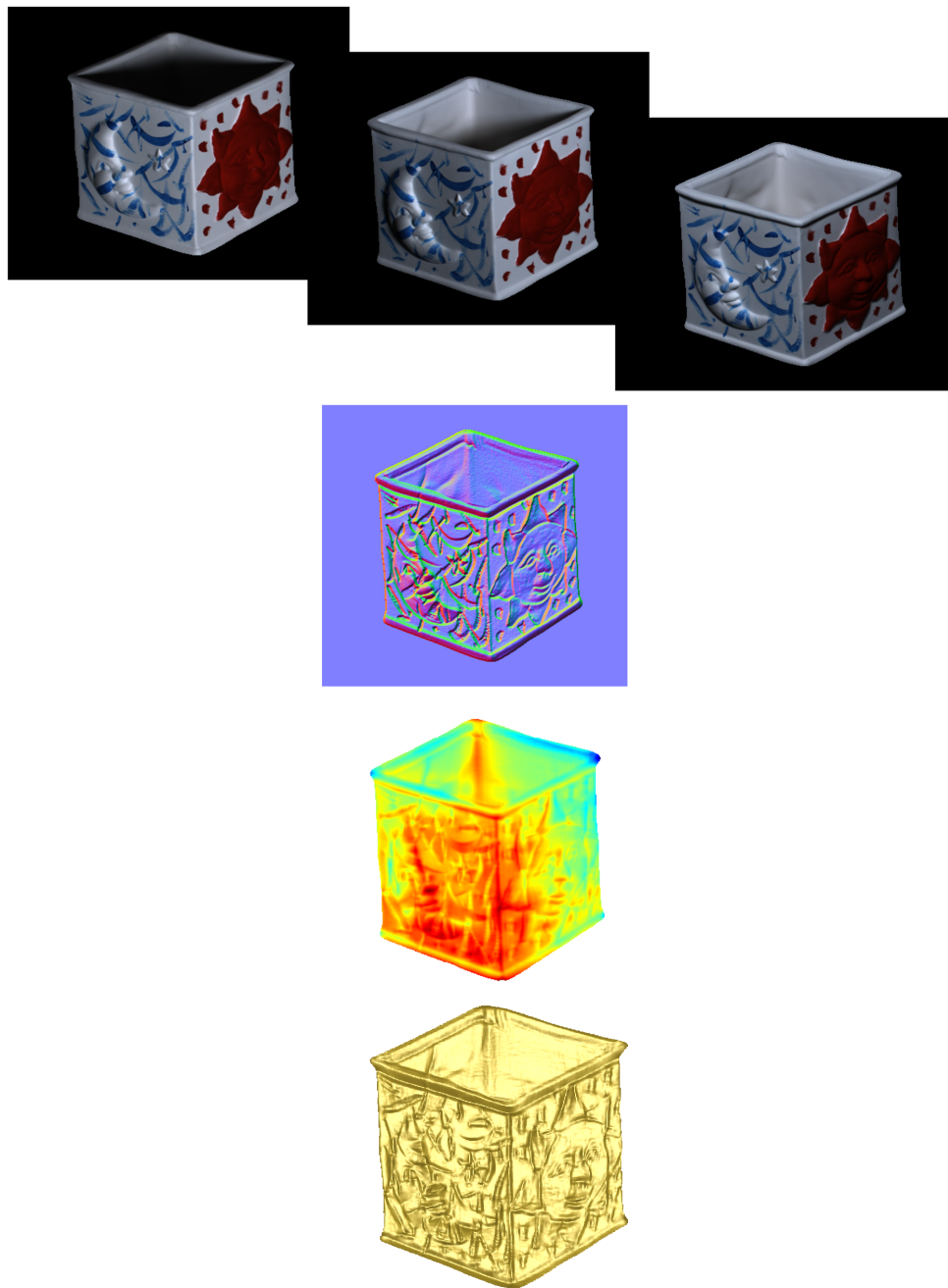


Figure 6-9: First row: three input real-world images from MIT Intrinsic Images [55]. Second, third and last rows: our normal, depth map and 3D reconstruction obtained by the proposed PS model. We think that our approach is successful in producing faithful normals, depth and 3D reconstruction for uncalibrated PS.

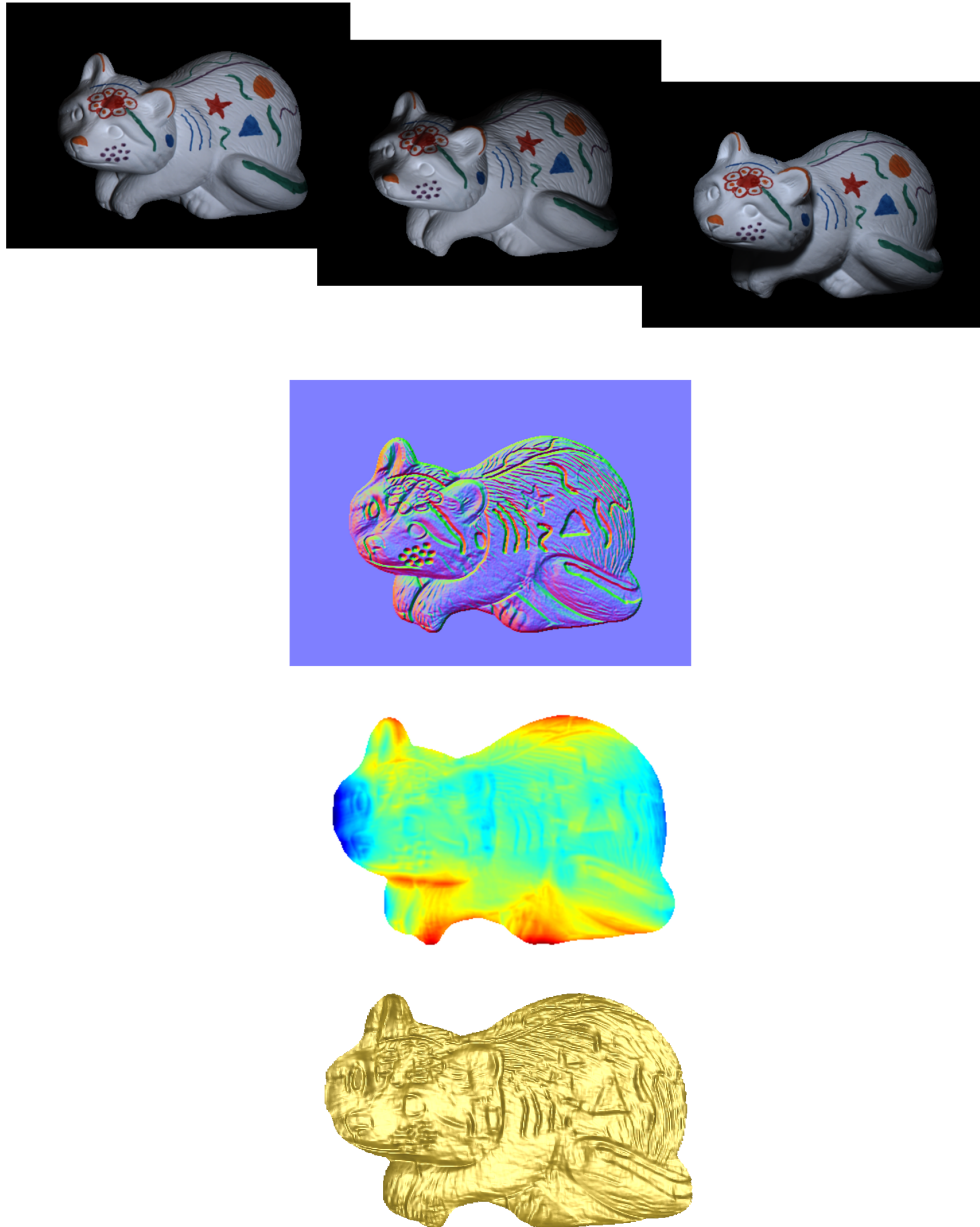


Figure 6-10: First row: three input real-world images from MIT Intrinsic Images [55]. Second, third and last rows: our normal, depth map and 3D reconstruction obtained by the proposed PS model. Our model could reconstruct all these information simultaneously as well.

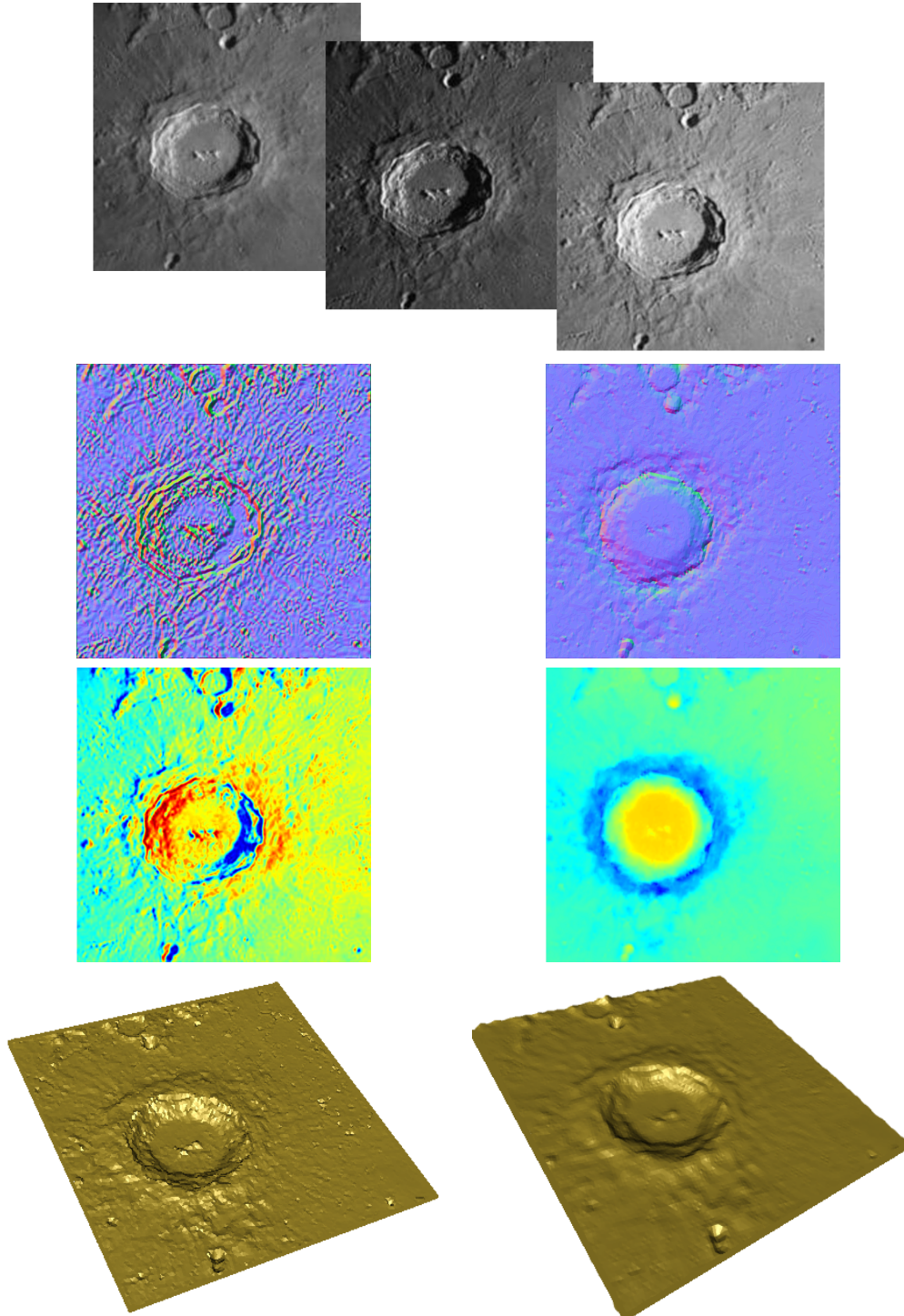


Figure 6-11: First row: three input internet real-world images from Moon [2]. Second, third and last rows: our normal, depth map and 3D reconstruction obtained by the proposed PS model (left) and by the deep network (right). It can be seen that applying meta heuristic PS on deep network outputs can improve results specially for tiny details. These results obtained without any information about the light direction, scene, etc.. Astronomical images can be regarded as a challenging experiment, our technique is able to provide reasonable results even for this experiment.

Chapter 7

Conclusion

7.1 Summary

Photometric stereo is one of the important 3D shape estimation approaches. By probing the shading variations under different illumination directions, photometric stereo is able to estimate the surface normal map. The obtained normal map can be applied to reconstruct the 3D shape of surfaces. The applicability of photometric stereo is really restricted by the assumptions that the classic photometric stereo relies on, particularly the diffuse surfaces, distant lighting, orthographic projection and pre-knowledge about light direction assumptions. These confined assumptions can seldom be held in real world applications. This dissertation focuses on generalizing photometric stereo method to make it applicable for more diverse cases and practical applications. To this aim, we have proposed different solutions for general reflectance models, spatially varying light, a supervised learning approach and finally a deep meta heuristic method. Moreover, we have proposed the photometric stereo techniques that work in a wild setup assuming unknown and general environment lightings as well as uncontrolled environment. All these efforts greatly improve the applicability of photometric stereo to various real world applications, especially for reconstructing 3D surfaces with fine details.

7.2 Future directions

This dissertation is concluded by mentioning some future improvements that we believe are important to pursue. One interesting future work can be providing the complete-view 3D reconstruction. This can be done by combining the photometric stereo with other 3D reconstruction techniques for example, multi-view stereo. Another point that we would like to consider is offering 3D reconstruction on portable devices such as smart phones, tablets and so on. Moreover, one assumption considered in the photometric stereo is that the surface should be fixed. So it would be nice if we can extend photometric stereo to work for the moving objects.

Bibliography

- [1] Blinn-phong shading model. https://en.wikipedia.org/wiki/Blinn_Phong_shading_model. Accessed: 1.10.2018.
- [2] Copernicus lunar crater. <http://pestoonak.blogfa.com/>. Accessed: 21.1.2018.
- [3] Cuckoo eggs. <http://what-when-how.com/birds/reed-warbler-birds/>, <http://bigpictures.club/makenewimage.html>, http://www.thepinsta.com/cuckoo-s-vs-warbler-eggs_zW8wWIUC0rr0XVfdBzOquPjtPBHFxFE1ugDGbxdIie4/. Accessed: 21.07.2018.
- [4] Humayun miniature. <https://en.wikipedia.org/wiki/Humayun>. Accessed: 21.07.2018.
- [5] Illumination and shading (basic computer graphics) part 1. http://what_when_how.com/computer_graphics_and_geometric_modeling/illumination_and_shading_basic_computer_graphics_part_1/. Accessed: 21.1.2018.
- [6] The life of birds. <http://www.pbs.org/lifeofbirds/home/>. Accessed: 10.01.2018.
- [7] Mirror hall. <https://en.wikipedia.org/wiki/Kamal-ol-molk>. Accessed: 21.07.2018.
- [8] msr-zbethel-tu, github repository, d. maturana. <https://github.com/dimatura/msr-zbethel-tu/tree/master/MSR-Driver/Media/statue2.obj>. Accessed: 2016.
- [9] Nasir-ol-molk mosque. <https://augtellez.wordpress.com/images/#jp-carousel-15153>. Accessed: 21.07.2018.
- [10] The picture is publicly available via license: CC BY-NC 4.0. <https://creativecommons.org/licenses/by-nc/4.0/>. Accessed: 21.1.2018.

- [11] @roxlu, jadi, github repository. <https://github.com/roxlu/jadi/blob/master/demos/ambient-occlusion/bin/data/statue.obj>. Accessed: 2016.
- [12] @roxlu, jadi, github repository. <https://www.yobi3d.com/>. Accessed: 10.9.2017.
- [13] The stanford 3d scanning repository. <http://graphics.stanford.edu/data/3Dscanrep/>. Accessed: 21.1.2016.
- [14] Structure of a biological neuron. https://www.wpclipart.com/medical/anatomy/nervous_system/neuron/neuron.png.html. Accessed: 21.1.2018.
- [15] Structure of an artificial neuron. <https://mizzlrblog.wordpress.com/2016/06/29/ascii-net-a-neural-network-for-character-recognition/>. Accessed: 21.1.2018.
- [16] Zandiyeh complex. <http://www.irantravelingcenter.com/zandiyeh-complex-shiraz/>. Accessed: 21.07.2018.
- [17] M. Ashikmin and P. Shirley. An anisotropic phong light reflection model. *Univ. of Utah TR-UUCS-00-014, Salt Lake City, UT*, 2000.
- [18] J. M. Avilés, B. G. Stokke, A. Moksnes, E. Roskaft, M. Asmul, and A. P. Moller. Rapid increase in cuckoo egg matching in a recently parasitized reed warbler population. *Journal of evolutionary biology*, 19(6):1901–1910, 2006.
- [19] M. Bähr, M. Breuß, Y. Quéau, A. S. Boroujerdi, and J. D. Durou. Fast and accurate surface normal integration on non-rectangular domains. *Computational Visual Media*, 3(2):107–129, 2017.
- [20] H. H. Balci and J. F. Valenzuela. Scheduling electric power generators using particle swarm optimization combined with the lagrangian relaxation method. *International Journal of Applied Mathematics and Computer Science*, 14:411–421, 2004.
- [21] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003.
- [22] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.
- [23] P. Beckmann and A. Spizzichino. The scattering of electromagnetic waves from rough surfaces. *Proceedings of the IEEE*, 52(11):1389–1390, 1964.
- [24] S. Binitha, S. S. Sathya, et al. A survey of bio inspired optimization algorithms. *International Journal of Soft Computing and Engineering*, 2(2):137–151, 2012.

- [25] J. F. Blinn. Models of light reflection for computer synthesized. In *ACM SIGGRAPH Computer Graphics*, volume 11, pages 192–198, 1977.
- [26] M. A. Boden. *Mind as machine: A history of cognitive science*. Clarendon Press, 2006.
- [27] B. Bontoux and D. Feillet. Ant colony optimization for the traveling purchaser problem. *Computers & Operations Research*, 35(2):628–637, 2008.
- [28] W. M. Brandenburg and J. T. Neu. Unidirectional reflectance of imperfectly diffuse surfaces. *Journal of the Optical Society of America*, 56(1):97–103, 1966.
- [29] M. L. Braunstein and J. W. Payne. Perspective and form ratio as determinants of relative slant judgments. *Journal of Experimental Psychology*, 81(3):584, 1969.
- [30] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76–90, 1970.
- [31] H. M. Cartwright. *Applications of evolutionary computation in chemistry*, volume 110. Springer Science & Business Media, 2004.
- [32] M. Chandraker, J. Bai, and R. Ramamoorthi. On differential photometric reconstruction for unknown, isotropic brdfs. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2941–2955, 2013.
- [33] M. Chandraker and R. Ramamoorthi. What an image reveals about material reflectance. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1076–1083. IEEE, 2011.
- [34] V. Chellaboina and M. K. Ranga. Reduced order optimal control using genetic algorithms. In *American Control Conference, 2005. Proceedings of the 2005*, pages 1407–1412. IEEE, 2005.
- [35] W. C. Cheng. Neural-network-based photometric stereo for 3d surface reconstruction. In *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pages 404–410. IEEE, 2006.
- [36] S. Y. Cho and T. W. S. Chow. Shape recovery from shading by a new neural-based reflectance model. *IEEE Transactions on Neural Networks*, 10(6):1536–1541, 1999.
- [37] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24, 1982.
- [38] M. Daoudi, A. Srivastava, and R. Veltkamp. *3D face modeling, analysis and recognition*. John Wiley & Sons, 2013.
- [39] N. B. Davies. Cuckoo adaptations: trickery and tuning. *Journal of Zoology*, 284(1):1–14, 2011.

- [40] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271, 1959.
- [41] M. Dorigo, G. D. Caro, and L. M. Gambardella. Ant algorithms for discrete optimization. *Artificial life*, 5(2):137–172, 1999.
- [42] D. Edwards, S. Boulos, J. Johnson, P. Shirley, M. Ashikhmin, M. Stark, and C. Wyman. The halfway vector disk for brdf modeling. *ACM Transactions on Graphics (TOG)*, 25(1):1–18, 2006.
- [43] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2650–2658, 2015.
- [44] A. P. Engelbrecht. *Fundamentals of computational swarm intelligence*. John Wiley & Sons, 2006.
- [45] A. R. Farooq, M. L. Smith, L. N. Smith, and S. Midha. Dynamic photometric stereo for on line quality control of ceramic tiles. *Computers in industry*, 56(8-9):918–934, 2005.
- [46] R. Fletcher. A new approach to variable metric algorithms. *The computer journal*, 13(3):317–322, 1970.
- [47] D. B. Fogel. An evolutionary approach to the traveling salesman problem. *Biological Cybernetics*, 60(2):139–144, 1988.
- [48] M. Galo and C. L. Tozzi. Surface reconstruction using multiple light sources and perspective projection. In *International Conference on Image Processing*, volume 2, pages 309–312, 1996.
- [49] Z. W. Geem, J. H. Kim, and G. V. Loganathan. A new heuristic optimization algorithm: harmony search. *simulation*, 76(2):60–68, 2001.
- [50] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence*, 23(6):643–660, 2001.
- [51] A. S. Glassner. *Principles of digital image synthesis: Vol. 1*, volume 1. Elsevier, 1995.
- [52] F. Glover. A template for scatter search and path relinking. In *European Conference on Artificial Evolution*, pages 1–51. Springer, 1997.
- [53] D. Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of computation*, 24(109):23–26, 1970.
- [54] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *CVPR*, volume 1, pages 341–348, 2005.

- [55] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, pages 2335–2342, 2009.
- [56] C. Gullon. *Height recovery of rough surfaces from intensity images*. PhD thesis, Heriot-Watt University, 2003.
- [57] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2004.
- [58] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 11(11):3079–3089, 1994.
- [59] K. He, X. Zhang, and J. Ren, S. and Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [60] A. Hertzmann and S. M. Seitz. Shape and materials by example: a photometric stereo approach. In *CVPR*, volume 1, pages 533–540, 2003.
- [61] N. Hoffman. Crafting physically motivated shading models for game development. part of “*Physically Based Shading Models in Film and Game Production*,” *SIGGRAPH*, 2010.
- [62] J. H. Holland. Outline for a logical theory of adaptive systems. *Journal of the ACM (JACM)*, 9(3):297–314, 1962.
- [63] B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. 1970.
- [64] B. K. P. Horn. *Robot Vision*. The M.I.T. Press, 1986.
- [65] B. K. P. Horn, R. J. Woodham, and W. M. Silver. *Determining shape and reflectance using multiple images*. M.I.T. Artificial Intelligence Laboratory, Memo 490, 1978.
- [66] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich. State of the art in transparent and specular object reconstruction. In *EUROGRAPHICS 2008 STAR–STATE OF THE ART REPORT*. Citeseer, 2008.
- [67] S. Ikehata and K. Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2179–2186, 2014.
- [68] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa. Robust photometric stereo using sparse regression. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 318–325. IEEE, 2012.
- [69] M. K. Johnson and E. H. Adelson. Retrographic sensing for the measurement of surface texture and shape. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1070–1077. IEEE, 2009.

- [70] D. Y. Kang, Y. J. Jang, and S. Won. Development of an inspection system for planar steel surface using multispectral photometric stereo. *Optical Engineering*, 52(3):039701, 2013.
- [71] D. Karaboga. An idea based on honey bee swarm for numerical optimization. Technical report, Technical report-tr06, Erciyes university, engineering faculty, computer engineering department, 2005.
- [72] J. Kennedy and R. C. Eberhart. The particle swarm: social adaptation in information-processing systems. In *New ideas in optimization*, pages 379–388. McGraw-Hill Ltd., UK, 1999.
- [73] M. Khanian. *Self depicted figures and self taken photographs*. 2018.
- [74] M. Khanian, A. Sharifi Boroujerdi, and M. Breuß. Perspective photometric stereo beyond lambert. In *QCAV*, volume 9534, pages 95341F–95341F–8, 2015.
- [75] M. Khanian, A. Sharifi Boroujerdi, and M. Breuß. Photometric stereo for strong specular highlights. *Computational Visual Media*, 2017, Springer, accepted for publication, 2017.
- [76] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.
- [77] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [78] M. Kurt and D. Edwards. A survey of brdf models for computer graphics. *ACM SIGGRAPH Computer Graphics*, 43(2):4, 2009.
- [79] J. Lambert and D. L. DiLaura. Photometry, or, on the measure and gradations of light, colors, and shade: Translation from the latin of photometria, sive, de mensura et gradibus luminis, colorum et umbrae. *Illuminating Engineering Society of North America*, 2001.
- [80] K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 5:164–168, 1944.
- [81] A. S. Lewis and M. L. Overton. Nonsmooth optimization via bfgs. *Submitted to SIAM J. Optimiz*, pages 1–35, 2009.
- [82] M. A. Lewis, P. K. Maini, and S. V. Petrovskii. Dispersal, individual movement and spatial ecology. *Lecture Notes in Mathematics–Mathematics Bioscience Series*, 2071, 2013.
- [83] C. T. Lin, W. C. Cheng, and S. F. Liang. Neural-network-based adaptive hybrid-reflectance model for 3-d surface reconstruction. *IEEE transactions on neural networks*, 16(6):1601–1615, 2005.

- [84] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato. Uncalibrated photometric stereo for unknown isotropic reflectances. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1490–1497. IEEE, 2013.
- [85] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An invitation to 3d vision: From images to models* springer verlag, 2003.
- [86] S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond lambert: reconstructing specular surfaces using color. In *CVPR*, volume 2, pages 619–626, 2005.
- [87] D. Marquardt. An algorithm for least squares estimation on nonlinear parameters. *Journal of the Society of Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [88] G. McGunnigle. *The classification of textured surfaces under varying illuminant direction*. PhD thesis, Citeseer, 1998.
- [89] R. Mecca, E. Rodola, and D. Cremers. Realistic photometric stereo using partial differential irradiance equation ratios. *Computers and Graphics*, 51:8–16, 2015.
- [90] R. Mecca, A. Tankus, and A. F. Bruckstein. Two-image perspective photometric stereo using shape-from-shading. In *ACCV*, volume 7727, pages 110–121, 2012.
- [91] D. Miyazaki, K. Hara, and K. Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision*, 86(2):229–242, 2010.
- [92] Y. Mukaigawa, Y. Ishii, and T. Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Society of America*, 24(10):3326–3334, 2007.
- [93] S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: physical and geometrical perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 13(7):611–634, 1991.
- [94] A. Ngan, F. Durand, and W. Matusik. Experimental analysis of brdf models. *Rendering Techniques*, 2005(16th):2, 2005.
- [95] H. B. Nielsen and K. Madsen. Introduction to optimization and data fitting. *month*, page 176, 2010.
- [96] J. F. Norman, J. T. Todd, H. F. Norman, A. M. Clayton, and T. R. McBride. Visual discrimination of local surface structure: Slant, tilt, and curvedness. *Vision research*, 46(6):1057–1069, 2006.
- [97] P. S. Ogun, M. R. Jackson, and R. M. Parkin. In-process surface profile assessment of rotary machined timber using a dynamic photometric stereo technique. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 226(6):823–830, 2012.

- [98] G. Oxholm and K. Nishino. Multiview shape and reflectance from natural illumination. In *CVPR*, pages 2163–2170, 2014.
- [99] T. Papadimitri and P. Favaro. A new perspective on uncalibrated photometric stereo. In *CVPR*, pages 1474–1481, 2013.
- [100] V. Parot, D. Lim, G. González, G. Traverso, N. S. Nishioka, B. J. Vakoc, and N. J. Durr. Photometric stereo endoscopy. *Journal of biomedical optics*, 18(7):076017, 2013.
- [101] M. Pharr and G. Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc, 2010.
- [102] B. T. Phong. Illumination for computer generated pictures. *Communications of ACM* 18, 18(6):311–317, 1975.
- [103] M. D. Plumbley. Algorithms for nonnegative independent component analysis. *IEEE Transactions on Neural Networks*, 14(3):534–543, 2003.
- [104] M. J. D. Powell. A hybrid method for nonlinear equations. *Numerical methods for nonlinear algebraic equations*, 1970.
- [105] L. Quan, H. Y. Shum, et al. Highlight removal by illumination-constrained inpainting. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 164–169. IEEE, 2003.
- [106] Y. Quéau and J. D. Durou. Edge-preserving integration of a normal field: weighted least squares, tv and l1 approaches. In *SSVM*, volume 9087, pages 576–588, 2015.
- [107] I. Rechenberg. *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. 1970.
- [108] G. Riegler, A. O. Ulusoy, and A. Geiger. Octnet: Learning deep 3d representations at high resolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 3, 2017.
- [109] A. Rosenberg, N. J. Cowan, and D. E. Angelaki. The visual representation of 3d object orientation in parietal cortex. *Journal of Neuroscience*, 33(49):19352–19361, 2013.
- [110] S. Rusinkiewicz. A survey of brdf representation for computer graphics. *CS348c. Universidad de Standford*, 1997.
- [111] B. Sathyabama, S. Raju, and A. Varadhan. Quaternion photometric stereo for rotation invariant surface texture classification. *American Journal of Applied Sciences*, 8(10):992, 2011.
- [112] J. A. Saunders and D. C. Knill. Perception of 3d surface orientation from skew symmetry. *Vision research*, 41(24):3163–3183, 2001.

- [113] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 519–528. IEEE, 2006.
- [114] S. A. Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.
- [115] D. F. Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of computation*, 24(111):647–656, 1970.
- [116] H. L. Shen and Q. Y. Cai. Simple and efficient method for specular removal in an image. *Applied optics*, 48(14):2711–2719, 2009.
- [117] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1850–1857. IEEE, 2009.
- [118] P. Shirley. Fundamentals of computer graphics. *MA: AK Peters Wellesley*, pages 301–346, 2005.
- [119] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE transactions on pattern analysis and machine intelligence*, 12(5):435–446, 1990.
- [120] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [121] M. L. Smith. The analysis of surface texture using photometric stereo acquisition and gradient space domain mapping. *Image and vision computing*, 17(14):1009–1019, 1999.
- [122] W. A. P. Smith and F. Fang. Height from photometric ratio with model-based light source selection. *Computer Vision and Image Understanding*, 145:128–138, 2016.
- [123] R. Socher, B. Huval, B. Bath, C. D. Manning, and A. Y. Ng. Convolutional-recursive deep learning for 3d object classification. In *Advances in Neural Information Processing Systems*, pages 656–664, 2012.
- [124] K. A. Stevens. Surface tilt (the direction of slant): A neglected psychophysical variable. *Attention, Perception, & Psychophysics*, 33(3):241–250, 1983.
- [125] H. Sugihara, I. Murakami, K. V. Shenoy, R. A. Andersen, and H. Komatsu. Response of mstd neurons to simulated 3d orientation of rotating planes. *Journal of Neurophysiology*, 87(1):273–285, 2002.
- [126] R. W. Sumner and J. Popović. Deformation transfer for triangle meshes. In *ACM SIGGRAPH*, volume 4, pages 399–405, 2004.

- [127] C. Szegedy et al. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [128] R. Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [129] H. D. Tagare and R. J. P. Defigueiredo. A framework for the construction of general reflectance maps for machine vision. *CVGIP: Image Understanding*, 57(3):265–282, 1993.
- [130] R. T. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE transactions on pattern analysis and machine intelligence*, 27(2):178–193, 2005.
- [131] K. L. Tang, C. K. Tang, and T. T. Wong. Dense photometric stereo using tensorial belief propagation. In *CVPR*, volume 1, pages 132–139, 2005.
- [132] Y. Tang, R. Salakhutdinov, and G. Hinton. Deep lambertian networks. *arXiv preprint arXiv:1206.6445*, 2012.
- [133] A. Tankus and N. Kiryati. Photometric stereo under perspective projection. In *ICCV*, volume 1, pages 611–616, 2005.
- [134] A. Tankus, N. Sochen, and Y. Yeshurun. Reconstruction of medical images by perspective shape-from-shading. In *ICPR*, pages 778–781, 2004.
- [135] A. Tankus, N. Sochen, and Y. Yeshurun. Shape-from-shading under perspective projection. *Computers and Graphics*, 63(1):21–43, 2005.
- [136] K. Tatemasu, Y. Iwahori, T. Nakamura, S. Fukui, R. J. Woodham, and K. Kasugai. Shape from endoscope image based on photometric and geometric constraints. *Procedia Computer Science*, 22:1285–1293, 2013.
- [137] S. Tibau, B. Willems, E. Van Den Bergh, and J. Wagemans. The role of the centre of projection in the estimation of slant from texture of planar surfaces. *Perception*, 30(2):185–193, 2001.
- [138] N. Tsumura, Y. Miyake, and F. H. Imai. Medical vision: measurement of skin absolute spectral-reflectance image and the application to component analysis. In *Proceedings of the 3rd International Conference on Multispectral Color Science (MCS'01)*, pages 25–28, 2001.
- [139] J. Unger, A. Wenger, T. Hawkins, A. Gardner, and P. Debevec. Capturing and rendering with incident light fields. Technical report, UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY CA INST FOR CREATIVE TECHNOLOGIES, 2003.
- [140] H. A. Varol and Z. Bingul. A new pid tuning technique using ant algorithm. In *American Control Conference, 2004. Proceedings of the 2004*, volume 3, pages 2154–2159. IEEE, 2004.

- [141] F. Verbiest and L. Van Gool. Photometric stereo with coherent outlier handling and confidence estimation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [142] O. Vogel, L. Valgaerts, M. Breuß, and J. Weickert. Making shape from shading work for real-world images. In *DAGM Pattern Recognition*, volume 5748, pages 191–200. Springer Berlin Heidelberg, 2009.
- [143] L. B. Wolff. Relative brightness of specular and diffuse reflection. *Optical Engineering*, 33(1):285–294, 1994.
- [144] R. J. Woodham. Photometric stereo: a reflectance map technique for determining surface orientation from image intensity. In *Image Understanding Systems and Industrial Applications, SPIE*, volume 0155, pages 136–143, 1978.
- [145] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):134–144, 1980.
- [146] C. C. Wu, K. C. Lai, and R. Y. Sun. Ga-based job scheduling strategies for fault tolerant grid systems. In *Asia-Pacific Services Computing Conference, 2008. APSCC'08. IEEE*, pages 27–32. IEEE, 2008.
- [147] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *ACCV*, volume 6494, pages 703–717, 2010.
- [148] T. P. Wu and C. K. Tang. Photometric stereo via expectation maximization. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):546–560, 2010.
- [149] Q. Yang, S. Wang, and N. Ahuja. Real-time specular highlight removal using bilateral filtering. In *European conference on computer vision*, pages 87–100. Springer, 2010.
- [150] X. S. Yang and S. Deb. Cuckoo search via lévy flights. In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, pages 210–214. IEEE, 2009.
- [151] C. Yu, Y. Seo, and S. W. Lee. Photometric stereo from maximum feasible lambertian reflections. In *European Conference on Computer Vision*, pages 115–126. Springer, 2010.
- [152] C. Yu, Y. Seo, and S.W. Lee. Photometric stereo from maximum feasible lambertian reflections. In *ECCV*, volume 6314, pages 115–126, 2010.
- [153] S. Zafeiriou, G. A. Atkinson, M. F. Hansen, W. A. P. Smith, V. Argyriou, M. Petrou, M. L. Smith, and L. N. Smith. Face recognition and verification using photometric stereo: The photoface database and a comprehensive evaluation. *IEEE Transactions on Information Forensics and Security*, 8(1):121–135, 2013.

- [154] J. Zheng, K. Fraser, J. C. Ho, J. V. Mendez, E. Natalia, P. Mandal, C. C. Huang, and R. Chiou. E-quality control in solar panel inspection using photometric stereo reconstruction and 3d measurement. In *IIE Annual Conference. Proceedings*, page 3430. Institute of Industrial and Systems Engineers (IISE), 2013.
- [155] S. K. Zhou, G. Aggarwal, R. Chellappa, and D. W. Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):230–245, 2007.
- [156] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur. Color subspaces as photometric invariants. *International Journal of Computer Vision*, 79(1):13–30, 2008.