Material Recognition Meets 3D Reconstruction: Novel Tools for Efficient, Automatic Acquisition Systems

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Dipl.-Ing. Michael Weinmann

aus Karlsruhe

Bonn, Dezember 2015

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Erstgutachter:	Prof. Dr. Reinhard Klein
Zweitgutachter:	Prof. Dr.-Ing. Michael Goesele

Tag der Promotion:	22. Juli 2016
Erscheinungsjahr:	2017

Contents

ZUSAMMENFASSUNG

Die hochgenaue Erfassung von Geometrie- und Reflektanzeigenschaften von Objekten stellt seit Jahrzehnten eines der Hauptziele in den Bereichen der Computer Vision und der Computergrafik dar und kommt in zahlreichen Anwendungen in der Industrie sowie im Bereich des Kulturerbes zum Einsatz. Die Reproduktion feiner Strukturen in der Oberflächengeometrie sowie im Reflektanzverhalten ist insbesondere in den Bereichen des Visual Prototypings, der Werbeindustrie sowie der digitalen Erhaltung von Objekten eine allgegenwärtige Bedingung geworden. Allerdings sind die heutigen Digitalisierungsmethoden üblicherweise nur für einen sehr eingeschränkten Bereich bezüglich der Objektmaterialien anwendbar. Zudem mangelt es an hochpräzisen Verfahren zur Erfassung von Objekten mit komplexem Reflektionsverhalten jenseits von diffuser Reflektanz. Weiterhin drängt die Nachfrage bezüglich einer massenhaften Digitalisierung von Objekten immer mehr auf vollautomatische, hocheffiziente Digitalisierungsmethoden, die eine genaue Erfassung von Objekten in einer möglichst kurzen Zeit erlauben.

Diese Dissertation ist der Erforschung grundlegender Komponenten gewidmet, welche für einen effizienten, automatischen Digitalisierungsprozess von großer Bedeutung sind. Eine solche effiziente, vollautomatische Digitalisierung kann erfolgen, wenn die beiden bisher üblicherweise getrennt betrachteten Bereiche der Materialerkennung und der 3D-Rekonstruktion zusammengeführt werden. Diesbezüglich verdeutlicht diese Arbeit, dass eine zuverlässige Materialerkennung für die betrachteten Objekte eine effizientere Geometrieerfassung erlaubt. Daher liegen die Hauptziele dieser Dissertation in der Entwicklung neuer, robuster Geometrieerfassungstechniken für Oberflächen mit komplexem Reflektanzverhalten jenseits von diffuser Reflektanz sowie der Entwicklung robuster Techniken für die Materialerkennung.

Im Bereich der Geometrieerfassung zeigt diese Arbeit auf, dass Verfahren basierend auf der Beleuchtung von Objekten mit strukturiertem Licht, welche für die Erfassung von Objekten aus Materialien mit diffuser Reflektanz bis zu spekularer Reflektanz mit einer gewissen diffusen Reflektanzkomponente geeignet sind, in einem Aufnahmesetup mit mehreren Projektoren und Kameras bezüglich der Auflösung durch die Überlagerung von aus verschiedenen Ansichten projizierten

Mustern deutlich verbessert werden können. Da die Ergebnisse solcher triangulationsbasierter Geometrieerfassungsmethoden üblicherweise hochfrequentes Rauschen aufgrund von ungenau lokalisierten Korrespondenzen in aus verschiedenen Ansichten aufgenommenen Bildern sowie ungenauer Kalibration enthalten, wird zudem ein neues Verfahren vorgestellt, bei welchem die Geometrieerfassung basierend auf der Beleuchtung mit strukturiertem Licht mit komplementären Informationen bezüglich photometrischer Normalen ergänzt wird und somit deutlich genauere Rekonstruktionen ermöglicht werden. Zusätzlich stellt diese Dissertation eine neue, robuste Methode zur Erfassung der Geometrie von spiegelnden Objekten mit komplexer Oberflächengeometrie vor.

Diese Untersuchungen im Bereich der 3D-Rekonstruktion werden durch die Entwicklung neuer Methoden für eine zuverlässige Materialerkennung ergänzt, welche in einem initialen Schritt verwendet werden können, um die gegebenen Oberflächenmaterialien zu erkennen und somit eine effiziente Auswahl geeigneter, anschließend durchgeführter Erfassungsmethoden ermöglichen. Im Rahmen dieser Dissertation erfolgt die Betrachtung einer robusten Materialerkennung für Szenarien mit kontrollierter Umgebungsbeleuchtung, wie sie in speziellen Laborumgebungen erzeugt werden kann, sowie für Szenarien mit natürlicher Beleuchtung, die im Alltag gegeben ist.

Abschließend werden neue Konzepte hinsichtlich einer effizienten, vollautomatischen Erfassung von Objekten auf Basis der Techniken, die im Rahmen dieser Dissertation entwickelt wurden, diskutiert.

ABSTRACT

For decades, the accurate acquisition of geometry and reflectance properties has represented one of the major objectives in computer vision and computer graphics with many applications in industry, entertainment and cultural heritage. Reproducing even the finest details of surface geometry and surface reflectance has become a ubiquitous prerequisite in visual prototyping, advertisement or digital preservation of objects. However, today's acquisition methods are typically designed for only a rather small range of material types. Furthermore, there is still a lack of accurate reconstruction methods for objects with a more complex surface reflectance behavior beyond diffuse reflectance. In addition to accurate acquisition techniques, the demand for creating large quantities of digital contents also pushes the focus towards fully automatic and highly efficient solutions that allow for masses of objects to be acquired as fast as possible.

This thesis is dedicated to the investigation of basic components that allow an efficient, automatic acquisition process. We argue that such an efficient, automatic acquisition can be realized when material recognition "meets" 3D reconstruction and we will demonstrate that reliably recognizing the materials of the considered object allows a more efficient geometry acquisition. Therefore, the main objectives of this thesis are given by the development of novel, robust geometry acquisition techniques for surface materials beyond diffuse surface reflectance, and the development of novel, robust techniques for material recognition.

In the context of 3D geometry acquisition, we introduce an improvement of structured light systems, which are capable of robustly acquiring objects ranging from diffuse surface reflectance to even specular surface reflectance with a sufficient diffuse component. We demonstrate that the resolution of the reconstruction can be increased significantly for multi-camera, multi-projector structured light systems by using overlappings of patterns that have been projected under different projector poses. As the reconstructions obtained by applying such triangulation-based techniques still contain high-frequency noise due to inaccurately localized correspondences established for images acquired under different viewpoints, we furthermore introduce a novel geometry acquisition technique that complements the structured light system with additional photometric normals and results in significantly more accurate reconstructions. In addition, we also present a novel method to acquire the 3D shape of mirroring objects with complex surface geometry.

The aforementioned investigations on 3D reconstruction are accompanied by the development of novel tools for reliable material recognition which can be used in an initial step to recognize the present surface materials and, hence, to efficiently select the subsequently applied appropriate acquisition techniques based on these classified materials. In the scope of this thesis, we therefore focus on material recognition for scenarios with controlled illumination as given in lab environments as well as scenarios with natural illumination that are given in photographs of typical daily life scenes.

Finally, based on the techniques developed in this thesis, we provide novel concepts towards efficient, automatic acquisition systems.

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Dr. Reinhard Klein for accepting me in a highly competitive group in an interesting research area. He not only gave me the opportunity to work on interesting topics, but he also provided me with inspiration. I feel that I did learn a lot during my employment at his department.

Furthermore, I would like to thank Prof. Dr.-Ing. Michael Goesele for his interest in my work and his willingness to review my thesis despite his tremendous workload.

I am also very grateful to my co-authors Prof. Dr. Reinhard Klein, Christopher Schwartz, Roland Ruiters, Ralf Sarlette, Aljosa Osep, Prof. Dr. Jürgen Gall, Fabian Langguth, Prof. Dr.-Ing. Michael Goesele, Dennis den Brok, Julian Iseringhausen, Dr. Arno Zinke, Martin Rump, Rodrigo Martin, and Prof. Dr. Matthias B. Hullin for their support regarding publications that are part of this thesis and other publications achieved during my time as a PhD student at the University of Bonn that helped me to keep my motivation alive and develop a broader view on the respective research domain.

Furthermore, I really have to thank several people for giving helpful comments and advice regarding my work. In particular, I am grateful to Roland Ruiters, Christoph Peters, Christopher Schwartz and my brother Martin Weinmann for all the highly valuable discussions through all the years. Furthermore, I also want to thank Dennis den Brok, Raoul Wessel, Sebastian Merzbach and Prof. Dr. Thomas Schultz for technical discussions and Dr. Arno Zinke for his valuable advice.

Being involved in two European projects during the thesis, I also was given the chance to cooperate with other people with high reputation in related research domains and to learn a lot regarding the management of such projects. I greatly thank the European Commission for the funding I received in the scope of the two research projects of the European Community's Seventh Framework Programme (FP7/2007-2013): *3D-COFORM* (grant agreement n° 231809; 2008-2012) and *Harvest4D* (grant agreement n° 323567; 2013-2016), and the involved colleagues.

The lively discussions on a broad range of topics of everyday life with the persons I shared the office will also not be forgotten. In particular, I gratefully thank Julian

Iseringhausen, Christopher Schwartz, and Kai Schröder. Furthermore, I thank all the members of the Institute of Computer Graphics II at the University of Bonn for the individual discussions we had throughout the time.

Finally, most of all, I have to thank my parents Ursula and Robert, my brother Frank and his wife Tatsiana and my twin brother Martin for their great support during all the time.

LIST OF ABBREVIATIONS

ABRDF	Apparent BRDF
BRDF	Bidirectional Reflectance Distribution Function
BSSRDF	Bidirectional Scattering-Surface Reflectance Distribution Function
BTF	Bidirectional Texture Function
CPU	Central Processing Unit
DoG	Difference-of-Gaussian
GPU	Graphics Processing Unit
HDR	High Dynamic Range
HOG	Histogram of Oriented Gradients
IBR	Image-Based Rendering
IR	Infrared
LDR	Low Dynamic Range
LED	Light-Emitting Diode (lamp)
LoG	Laplacian-of-Gaussian
MR	Maximum Response
MSE	Mean Squared Error
MSER	Maximally Stable Extremal Regions
OpenGL	Open Graphics Library
PCA	Principal Component Analysis
QR	Quick Response
RAM	Random Access Memory
RANSAC	Random SAmple Consensus
RGB	Red Green Blue
RBF	Radial Basis Function
RMSE	Root Mean Squared Error
SBA	Sparse Bundle Adjustment
SIFT	Scale-Invariant Feature Transform
SVBRDF	Spatially Varying BRDF
SVD	Singular Value Decomposition
SVM	Support Vector Machine

Part I

Introduction

CHAPTER 1

INTRODUCTION

The rich information perceived via the senses of the human perceptual system such as sight, hearing, taste, smell, touch and balance greatly supports us in exploring our environment and, combined with our gained experience, allows us to infer insights regarding daily life tasks such as how we have to interact with the content of the surrounding environment. Among these senses, sight is probably the most important one for these interactions as it allows a touchless close-range and far-range perception of our environment, whereas the other senses allow a more limited interaction. The content we visually perceive in a scene is characterized by the presence and arrangement of objects, their shapes as well as their attached colors and textures. However, observed colors and textures are a result of the complexity of visual appearance due to the interplay of surface geometry, material properties and illumination conditions and, in turn, also affect the way we perceive 3D shapes. From the impression of the observed objects and materials, further insights regarding physical and functional properties such as their deformability, fragility, density, weight, value, thermal conductivity or toxicity can be derived. Based on examples of such phenomena that are not directly visible, the argumentation in [Fle14] concludes that there is evidence that impressions regarding materials might be learned associations. Indeed, based on visual perception, we not only get impressions about a characteristic look but also an accompanying “feel” for materials.

Consequently, as stated in one of the fragments attributed to the Greek philosopher Anaxagoras of Clazomenae (c. 500 B.C. – 428 B.C.),

> *appearances are a sight of the unseen* (Anaxagoras, fragment B 21),

i.e. sense perception allows to infer an understanding of an underlying more general concept which cannot be perceived based on appearances alone [Cur10].

Clearly, detecting, localizing and classifying the objects present in a local environment as well as reasoning about their 3D shapes and material properties are among the key tasks to be solved in daily life. However, it is also evident that exactly the same objectives merit an increasing attraction in industrial applications and that there is an urgent need for transferring the capabilities of the human visual system to fully automated systems.

1.1 Interrelating 3D Shape Acquisition and Material Recognition

Being one of the fundamental goals in the fields of computer vision and computer graphics for decades, the faithful reconstruction of 3D shape has many applications in quality assurance, reverse engineering, digital preservation in cultural heritage and entertainment. In particular, there is a demand for obtaining highly detailed and hole-free 3D surface geometries of objects, which is especially challenging for objects made of materials with complex non-Lambertian reflectance behavior. For the photo-realistic visual reproduction of real-world objects, reflectance characteristics have to be acquired as well, ideally directly on the true surface geometry. Automatic reflectance acquisition devices such as gonioreflectometers or camera arrays are capable of taking images of an object under a huge multitude of varying viewing and illumination conditions. In order to allow the reconstruction of the underlying surface geometry, these setups are typically equipped with laser scanners or structured light systems as well. A highly accurate acquisition of the object shape allows a subsequent, high-quality rendering of digitized objects. Consequently, there is a high demand for highly accurate geometry and reflectance acquisition techniques. However, accurately capturing optical properties of materials makes the acquisition process complicated for many materials, and there is still a need for acquisition techniques which are appropriate for handling surfaces with complex reflectance behavior such as mirroring surfaces, translucency, transparency, etc.. Unfortunately, current state-of-the-art automatic acquisition and reconstruction techniques are designed for only a limited range of surface reflectance and the user typically selects appropriate ones based on his experience (see Figure 1.1). This represents the typical acquisition scenario with respect to cultural heritage artifacts. In a similar way, the user has to select the respective industrial work flow depending on the material of the object in the scope of many industrial applications. In both cases, the ultimate prerequisite is formed by the existence of acquisition techniques appropriate for the different surface materials that might occur. For objects with heterogeneous surface reflectance behavior due to different surface materials, such as diffuse and mirroring parts of the surface, the acquisition expert has to select appropriate techniques for the different material types and to merge the resulting reconstructions (see Figure 1.2).

However, the demands for an automatic, robust and efficient acquisition process as

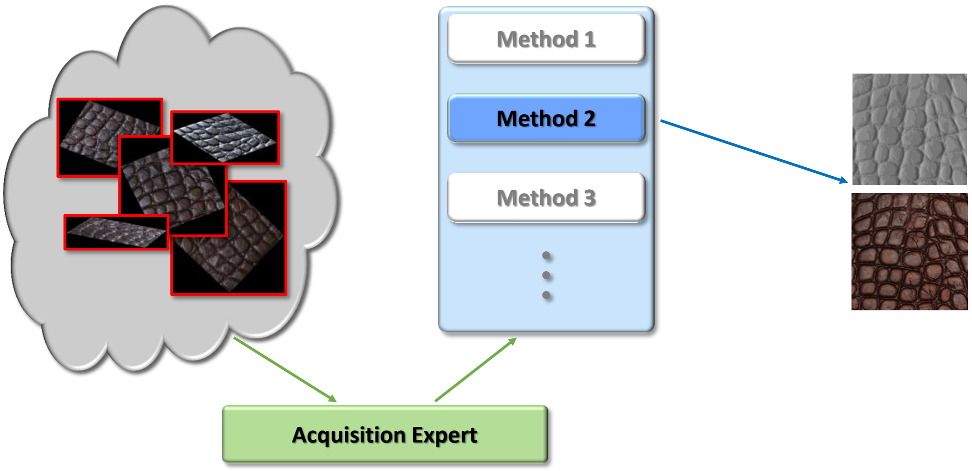


Figure 1.1: *User-guided acquisition process: An acquisition expert judges about the reflectance behavior of an object and selects appropriate acquisition techniques based on his experience.*

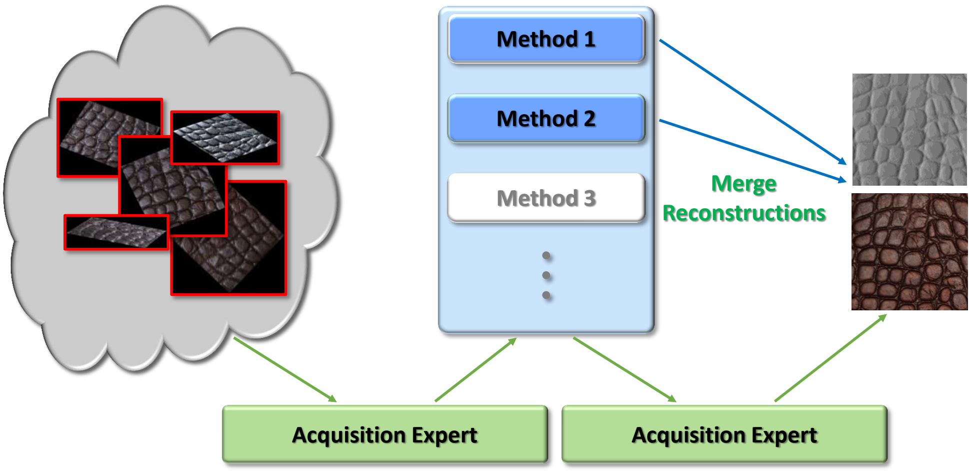


Figure 1.2: *User-guided acquisition process with manual merging step: An acquisition expert judges about the reflectance behavior of an object, selects appropriate acquisition techniques based on his experience and manually merges the results obtained from the individual techniques to a final reconstruction result.*

specified in industry require a shift in paradigms. Instead of the above-mentioned traditional user-guided acquisition schemes, the presence of the individual occurring surface materials should guide the acquisition process if there is no prior knowledge about the object surface available. Hence, automatically recognizing the occurring surface materials represents a key component for automatic acquisition pipelines. Indeed, this enables making decisions such as reasoning about fragility, deformability, weight, etc., which, in turn, naturally guide the interaction of humans with objects in daily life. For handling the individual objects appropriately in a supply chain using robotized control systems, work processes need to be adapted to their respective surface materials. Obviously, knowledge about materials and their locations on the object surface opens up the possibility for significantly increasing the efficiency of processes regarding acquisition time and resources. Instead of naïvely conducting acquisitions for different types of surface materials and merging the individual reconstructions as illustrated in Figure 1.3, knowledge regarding whether a surface region exhibits e.g. a rather Lambertian or a specular reflectance behavior allows to select appropriate 3D reconstruction techniques for each region on the object surface (see Figure 1.4). Furthermore, based on the material properties, there is also the possibility to automatically detect cases where none of the available reconstruction techniques is appropriate.

This thesis is dedicated to the development of techniques that serve as fundamental prerequisites for an efficient, automatic acquisition process. We demonstrate that such an efficient, automatic acquisition can be realized when material recognition "meets" 3D reconstruction, i.e. the task of acquiring 3D geometry can be approached more efficiently by reliably recognizing the respective surface materials of the involved object. In this regard, this thesis focuses on obtaining novel insights and developing novel tools for both improving state-of-the-art acquisition techniques for easy-to-handle surface materials and handling more complex surface materials. In particular, we present an extension of structured light systems, which, together with laser scanners, represent the standard regarding the acquisition of the 3D geometry of objects and are well-suited for objects ranging from diffuse surface reflectance to even specular surface reflectance with a sufficient diffuse component. While this extension is focused on increasing the resolution of multi-camera, multi-projector structured light systems, we additionally propose a novel geometry acquisition technique that complements the structured light system with additional photometric normals and results in significantly more accurate reconstructions. Furthermore, we introduce a novel method that allows the acquisition of the 3D shape of mirroring objects. These investigations are accompanied by the development of novel tools for reliable material recognition which can be used in an initial step to recognize the occurring surface materials and, hence, to efficiently select the subsequently applied, appropriate acquisition techniques based on these classified materials. In

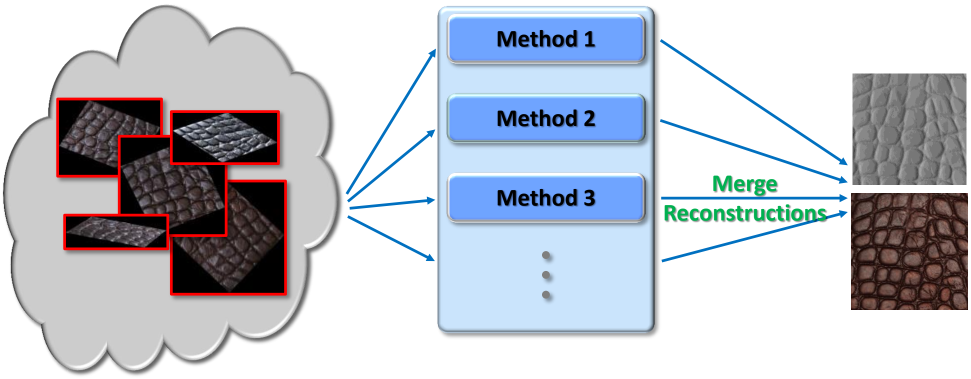


Figure 1.3: *Naïve automatic acquisition process: The object of interest is measured using all the available acquisition techniques. Subsequently, the individual reconstructions have to be merged in order to get an appropriate reconstruction.*

the scope of this thesis, we therefore consider material recognition under controlled illumination conditions as given in lab environments as well as material recognition under complex, real-world illumination conditions.

1.2 Challenges and Main Contributions

In the last decades, numerous investigations have focused on 3D reconstruction of objects. In general, active scanning techniques such as structured light systems or laser scanners have been proven to allow for a higher accuracy in the reconstruction result in comparison to passive techniques such as multi-view stereo approaches. However, facing the demand for an extremely high visual quality of digitized models with respect to geometric accuracy and photo-realism as e.g. required in cultural heritage, entertainment, etc., there is a need to even improve the accuracy of active scanning solutions to also capture fine surface details such as scratches or engravings which significantly contribute to the characteristic appearance of an object. In addition, there is evidence that materials with complex optical surface reflectance such as mirroring surfaces or translucent surfaces impose real challenges regarding their acquisition. Therefore, today's challenges definitively include the development of geometry acquisition techniques appropriate for this kind of surface materials as well.

Facing these challenges, this thesis advances the state-of-the-art regarding 3D reconstruction by introducing solutions for improving the accuracy of current scanning techniques and handling a wider range of different surface materials

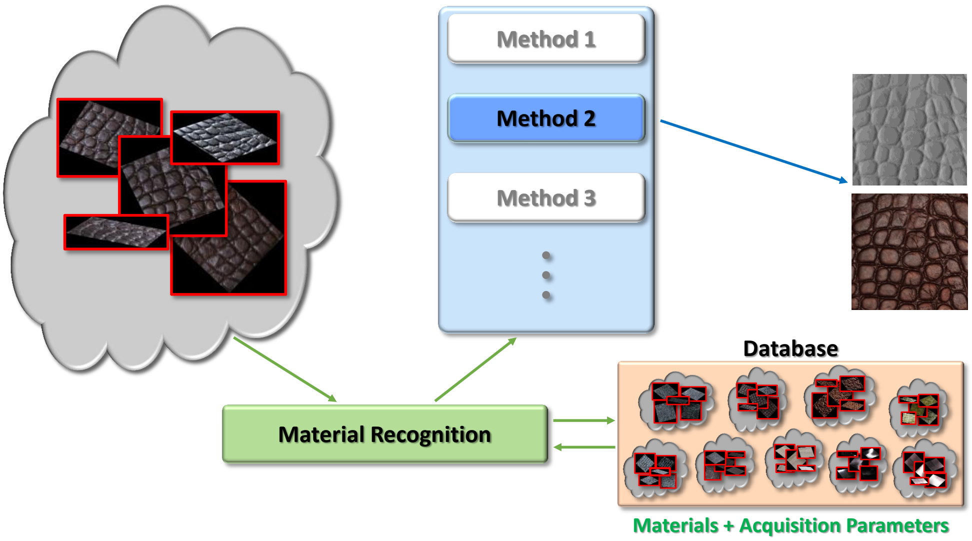


Figure 1.4: *Efficient, automatic acquisition process: Based on a prior material recognition step, the respective annotations of the most similar material in a database can be used to guide the acquisition process. Consequently, only the required techniques are involved, which leads to a significant increase in efficiency.*

beyond Lambertian surface reflectance. In particular, the key contributions in this regard are:

A novel super-resolution scheme to improve the accuracy of multi-camera, multi-projector systems based on structured light:
We have developed a structured light based multi-camera, multi-projector device which allows the reconstruction of the full 3D shape of objects without moving either the object or the acquisition setup. As a result, we circumvent the critical registration of independent measurements. So far, cameras still have considerably better resolutions in comparison to projectors. Therefore, the limiting factor in a multi-camera, multi-projector system can be identified in the resolution of the used projectors. To overcome this limitation, we have developed a novel super-resolution scheme which takes advantage of the fact that a sufficient number of projectors, or a projector sequentially mounted at several different positions, needs to be involved for covering the whole object surface with structured light patterns. By exploiting the fact that for almost all points on the surface the patterns of projectors at different positions overlap, much smaller regions on the surface can be uniquely identified. This allows to

reconstruct a much denser and more accurate point cloud. In addition, by using high dynamic range imaging, our technique is able to even handle complicated objects which exhibit strong specularities.

A novel approach to increase the accuracy of 3D reconstructions for opaque objects by fusing structured light consistency and Helmholtz normals:
By combining a structured light based consistency measure with dense normal information obtained by exploiting the Helmholtz reciprocity principle, our method overcomes the limitations of the individual techniques such as the low-frequency drift of techniques based on normal information or the high-frequency noise of triangulation-based techniques induced by inaccurately localized point correspondences or inaccuracies in calibration. The reconstruction is performed by solving one global variational problem which integrates all available measurements simultaneously, over all cameras, light source positions and rotation angles of the involved turntable. By employing an octree-based continuous min-cut framework, our technique alleviates metrification errors while still being memory-efficient.

A novel, robust multi-view normal field integration technique for 3D reconstruction of mirroring objects:
For mirroring objects, most traditional techniques such as laser scanners, structured light systems or multi-view stereo systems fail in providing reliable reconstructions. Therefore, several methods have been developed that make use of the observed information of specular highlights or reflections of the surrounding environment. However, none of the previous approaches has shown high-quality reconstructions for complex geometries in the presence of occlusions and interreflections. In this thesis, we have advanced the state-of-the-art in this regard by introducing a novel, robust multi-view normal field integration technique. In a turntable-based setup, displays are used to actively illuminate the mirroring object surface with patterns that are observed by cameras. This allows to calculate individual volumetric normal fields for each combination of camera, display and turntable angle. In order to be capable of also handling blurred pattern observations induced by surface curvature or imperfect mirroring surface characteristics, our technique takes care to locally adapt the decoding to the finest still resolvable pattern resolution. From the individual, overlapping normal fields, we infer for each point in the scene both the most likely local surface normal and a local surface consistency estimate via a non-parametric clustering of normal hypotheses. Subsequently, the estimates for the local surface normals and the local surface consistencies are taken as input to an iterative min-cut based variational 3D reconstruction approach.

Furthermore, the demand for as-efficient-as-possible solutions has become a crucial prerequisite in many industrial applications. Speeding up e.g. acquisition processes or quality inspection processes and reducing the required hardware capacities are key components for reducing the incurring total costs and, hence, have to be considered by companies when evaluating their competitiveness. In this regard, it is important to only spend the effort that is really needed for the respective task. Therefore, using highly efficient approaches is inevitable.

Focusing on the context of geometry and reflectance acquisition, this means that from a pool of material-specific acquisition techniques only those methods should be selected for which a certain part of the object surface exhibits the corresponding reflectance behavior assumed by these methods. For instance, to acquire the surface geometry of a heterogeneous object with both diffuse and mirroring surface parts, only a reliable shape acquisition technique for diffuse objects and a respective method for mirroring objects should be involved. In an initial stage before the actual acquisition, it is therefore desirable to reliably recognize the present surface materials. Subsequently, these recognized surface materials can be used to guide the acquisition process. As research in this context is still in an infant stage, this thesis aims at improving the state-of-the-art by providing tools for a reliable material recognition and their use within an efficient acquisition pipeline. In particular, the respective key contributions of this thesis towards image-based inference of surface materials are:

A novel technique for robust material recognition under controlled lighting conditions:
Most of the investigations in the related literature on appearance-based material recognition in images either focus on recognizing materials in single query images or on recognizing materials in a set of images depicting the same material sample under several registered view-light configurations which are also present in the training data. Using such a set of several images acquired under different view-light configurations provides significantly richer query information with a more detailed representation of characteristic material traits such as interreflections which might not necessarily appear under a single view-light configuration. As expected, this fact makes set-based material recognition typically more robust than single-image-based material recognition. However, different acquisition devices typically consider different view-light configurations, e.g. due to constraints with respect to placing the involved components, and, hence, make methods relying on training data and test data acquired under exactly the same registered view-light configurations impractical. Our technique overcomes these

limitations arising from the need for having material information from exactly the same view-light configurations in the same order available per material by establishing *material spaces*. These material spaces are computed per material from material characteristics observed in the images acquired under different view-light configurations via standard feature descriptors for color and texture, and a comparison of materials can be carried out directly on the respective material spaces via set-to-set distances. The robustness of our technique can be recognized in the fact that considering only a few view-light configurations already leads to high recognition rates.

A novel framework for recognizing materials under complex real-world illumination based on training data synthesized using a BTF database: Most of the approaches in literature focus on material recognition under controlled illumination conditions as given in lab environments and, typically, the material databases used in these investigations only consider a rather small subset of all the possible illumination conditions. As a consequence, these databases do not incorporate the complex variations in material appearance under natural illumination to be expected in most of our daily life situations and, hence, cannot serve as training data for material recognition under the respective illumination conditions. An alternative consists in manually acquiring different material samples in a multitude of locations with different illumination conditions under several viewing conditions. Unfortunately, the involved manual acquisition and the subsequently required manual annotation of the acquired images make this approach rather impractical. In order to bridge the gap and allow for material recognition under natural illumination, we propose utilizing the potential of computer graphics to generate synthetic training data from separately acquired material and illumination characteristics. For such data, segmentations regarding the involved surface materials can easily be derived which makes this approach particularly attractive for huge datasets. It will turn out that one key prerequisite for the success of this approach can be identified in an appropriate representation of material appearance which accurately captures the details of surface reflectance behavior. While BRDFs are only appropriate for smooth, homogeneous materials, many of the materials encountered in everyday life exhibit a more complex reflectance behavior. Focusing on such materials, we propose to use BTFs for generating synthetic data which depicts the digitized materials under a huge multitude of different view-light conditions. Our investigations will demonstrate that this synthesized training data allows for material recognition under complex natural illumination.

A novel framework for efficient, automatic acquisition of geometry and reflectance based on a guidance by an initial material recognition:
So far, acquisition processes are typically based on selecting appropriate techniques based on a-priori knowledge regarding the expected surface reflectance behavior, which is typically obtained from user experience. For a fully automatic acquisition under missing a-priori knowledge about the occurring surface materials, the combination of individual, material-specific acquisition techniques is inevitable. While naïvely applying techniques available for e.g. diffuse objects, opaque objects, mirroring objects, translucent objects or transparent objects respectively and merging the individual results in a subsequent stage would be possible, such a strategy might be too inefficient for many industrial applications where efficiency regarding time and hardware usage is required. Instead, we propose an efficient acquisition pipeline where only those acquisition techniques are involved that are really required for a considered material. For this purpose, our acquisition pipeline is based on a prior automatic material recognition step that allows to recognize the respective material itself or a very similar material in a material database with annotations regarding appropriate acquisition techniques for the incorporated materials. The evaluation of the accompanying annotations of the recognized material allows for an intelligent selection of the subset of techniques required during the acquisition process and, hence, allows for an efficient acquisition.

With our investigations, we were thus able to achieve substantial improvements in both fields of geometry reconstruction and material recognition and to even establish a connection between these two domains to increase the efficiency of the acquisition process which has, to the best of our knowledge, not been addressed in a similar way before.

1.3 Publications

Most of the material presented in this thesis has successfully passed a peer review. In particular, the respective publications are:

- M. Weinmann, C. Schwartz, R. Ruiters, and R. Klein:
  A Multi-Camera, Multi-Projector Super-Resolution Framework for Structured Light.
  In *Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pp. 397-404, 2011.
- M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein:
  Fusing Structured Light Consistency and Helmholtz Normals for 3D Recon-

struction.
In *Proceedings of the British Machine Vision Conference*, pp. 108.1-108.12, 2012.

- M. Weinmann, A. Osep, R. Ruiters, and R. Klein:
  Multi-View Normal Field Integration for 3D Reconstruction of Mirroring Objects.
  In *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 2504-2511, 2013.
- M. Weinmann, J. Gall, and R. Klein:
  Material Classification Based on Training Data Synthesized Using a BTF Database.
  In *Proceedings of the European Conference On Computer Vision (ECCV) - Part III*, pp. 156-171, 2014.
- M. Weinmann and R. Klein:
  Material Recognition for Efficient Acquisition of Geometry and Reflectance.
  In *Computer Vision - ECCV 2014 Workshops*, 8927, pp. 321-333, 2015.
- M. Weinmann and R. Klein:
  Advances in Geometry and Reflectance Acquisition.
  In *Proceedings of SIGGRAPH Asia 2015 Courses*, pp. 1:1-1:71, 2015.
- M. Weinmann, D. den Brok, S. Krumpen, and R. Klein:
  Appearance Capture and Modeling.
  In *Proceedings of SIGGRAPH Asia 2015 Courses*, pp. 4:1-4:1, 2015.
- M. Weinmann, F. Langguth, M. Goesele, and R. Klein:
  Advances in Geometry and Reflectance Acquisition.
  In *Eurographics 2016 Tutorials (accepted at)*, 2016.
- M. Weinmann, M. Weinmann, F. Rottensteiner, and B. Jutzi:
  Acquisition and Automatic Characterization of Scenes - From Point Clouds to Features and Objects.
  In *ISPRS Congress 2016 Tutorials (accepted at)*, 2016.

Furthermore, we included some of the discussions presented in this thesis in a positioning paper to provide a survey on the achievements and remaining challenges in the scope of material recognition:

- M. Weinmann and R. Klein:
  A Short Survey on Optical Material Recognition.
  In *Eurographics Workshop on Material Appearance Modeling*, pp. 35-42, 2015.

The structured light based geometry reconstruction and the automatic geometric calibration procedures presented in the scope of the papers included in this thesis have been proven to be a fundamental component with respect to improvements in the state-of-the-art in reflectance acquisition techniques which heavily rely on an accurate acquisition of surface geometry. Therefore, these components have also been used in several publications focusing on appearance acquisition which are not part of this thesis:

- C. Schwartz, M. Weinmann, R. Ruiters, and R. Klein:
  Integrated High-Quality Acquisition of Geometry and Appearance for Cultural Heritage.
  In *Proceedings of the International Symposium on Virtual Reality, Archeology and Cultural Heritage (VAST)*, pp. 25-32, 2011.
- C. Schwartz, R. Ruiters, M. Weinmann, and R. Klein:
  WebGL-Based Streaming and Presentation Framework for Bidirectional Texture Functions.
  In *Proceedings of the International Symposium on Virtual Reality, Archeology and Cultural Heritage (VAST)*, pp. 113-120, 2011.
- C. Schwartz, M. Weinmann, R. Ruiters, A. Zinke, R. Sarlette, and R. Klein:
  Capturing Shape and Reflectance of Food.
  In *Proceedings of SIGGRAPH Asia 2011 Sketches*, pp. 28:1-28:2, 2011.
- C. Schwartz, R. Sarlette, M. Weinmann, and R. Klein:
  DOME II: A Parallelized BTF Acquisition System.
  In *Proceedings of the Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition*, pp. 25-31, 2013.
- C. Schwartz, R. Ruiters, M. Weinmann, and R. Klein:
  WebGL-Based Streaming and Presentation of Objects With Bidirectional Texture Functions.
  In *Journal on Computing and Cultural Heritage (JOCCH)*, 6(3), pp. 11:1-11:21, 2013.
- C. Schwartz, R. Sarlette, M. Weinmann, M. Rump, and R. Klein:
  Design and Implementation of Practical Bidirectional Texture Function Measurement Devices Focusing on the Developments at the University of Bonn.
  In *Sensors*, 14(5), pp. 7753-7819, 2014.

Further closely related work with contributions from the author of this thesis has been published in:

- M. Weinmann, M. Weinmann, S. Hinz, and B. Jutzi:
  Fast and automatic image-based registration of TLS data.
  In *Proceedings of the ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (6), pp. 62-70, 2011.

- D. den Brok, M. Weinmann, and R. Klein:
  Linear Models for Material BTFs.
  In *Eurographics Workshop on Material Appearance Modeling*, pp. 15-19, 2015.
- R. Martin, J. Iseringhausen, M. Weinmann, and M. B. Hullin:
  Multimodal Perception of Material Properties.
  In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, pp. 33-40, 2015.

These contents, however, are not part of this thesis.

1.4 Thesis Outline

The thesis is organized in four parts. This introduction given in Part I provides a motivation for our work by introducing the related research domain and clarifying the relevance of the investigations performed in the scope of this thesis. Subsequently, an overview of the key contributions achieved in this thesis is given, which is followed by a list of peer-reviewed publications where most of the content of this thesis has been presented to the community. Additionally, this part contains a list of our publications that are not part of this thesis.

In Part II, we first provide an overview of material-specific acquisition of both geometry and reflectance in Chapter 2. This chapter reviews a grouping of surface materials with respect to the acquisition methods that can be applied to obtain appropriately reconstructed digital models as proposed in literature, and also provides an overview on techniques proposed in literature to handle the individual groups of surface materials. This is followed by a discussion of the contributions achieved in the context of 3D reconstruction which are represented by an improvement of conventional multi-camera, multi-projector structured light systems by using a novel super-resolution scheme (see Chapter 3), a further increase of the reconstruction quality by combining structured light information with normal information in the scope of a variational framework (see Chapter 4) and the development of a robust 3D reconstruction technique for mirroring objects in the presence of occlusions and interreflections due to complex surface geometry (see Chapter 5).

Part III then focuses on image-based inference of material characteristics (see Chapter 6) for which we briefly describe characteristic material attributes. Furthermore, we provide an overview of commonly used descriptors for capturing such material attributes from images as well as an overview of the main approaches followed with respect to material recognition. Subsequently, we discuss our contributions achieved in the scope of material recognition under controlled illumination

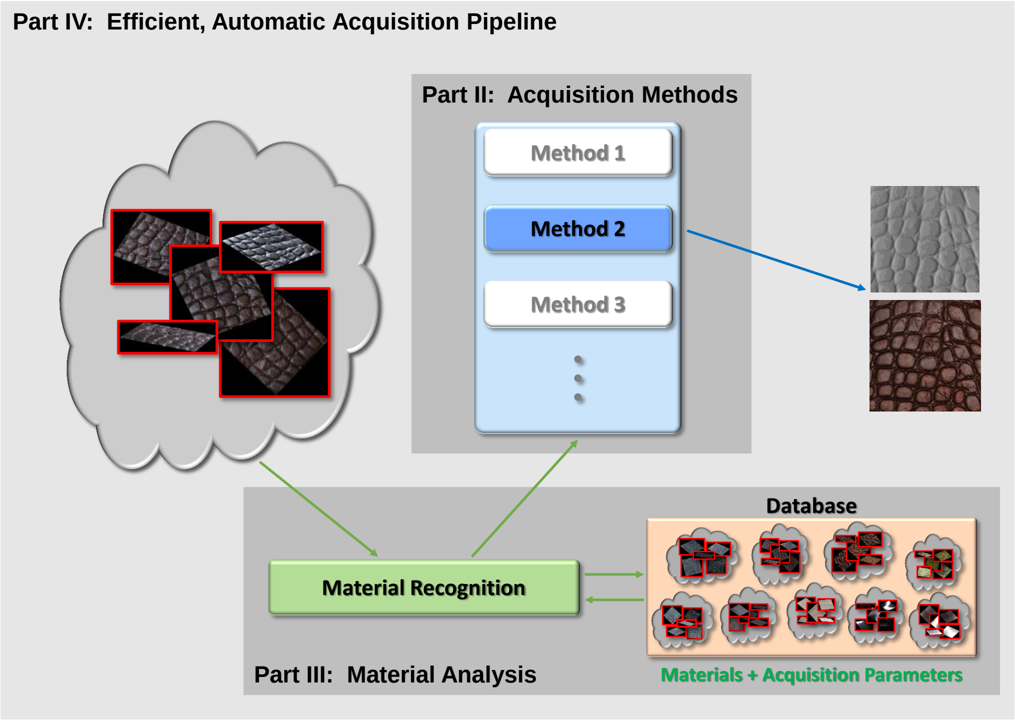

Part IV: Efficient, Automatic Acquisition Pipeline
Part II: Acquisition Methods
Method 1
Method 2
Method 3
Database
Material Recognition
Part III: Material Analysis
Materials + Acquisition Parameters

Figure 1.5: *Overview with respect to the structure of the technical parts of this thesis: In Part II, novel methods for the acquisition of objects with different surface materials are described. Part III focuses on the recognition of materials under both controlled and uncontrolled illumination conditions. Finally, Part IV is dedicated to the establishment of a connection between the contributions from the previous parts resulting in an efficient, automatic acquisition framework.*

conditions (see Chapter 7) and under more complex natural illumination conditions (see Chapter 8).

The final Part IV is dedicated to the establishment of the connection between material recognition and 3D reconstruction and introduces an efficient, automatic acquisition pipeline which allows an efficient acquisition process for both 3D geometry and reflectance (see Chapter 9). Subsequently, we summarize the contributions achieved in the scope of this thesis and provide a discussion of possibilities regarding future work (see Chapter 10).

The structure of the technical parts discussed in this thesis is illustrated in Figure 1.5.

Part II

A Survey on Advances in Acquisition and Our Novel Geometry Acquisition Methods for Different Complexities of Surface Materials

CHAPTER 2

A SURVEY ON MATERIAL-SPECIFIC ACQUISITION

The faithful digital reproduction of objects is an important topic with applications in areas such as industry, entertainment, marketing and cultural heritage. Important tasks can be identified in quality assurance, reverse engineering or digital preservation. In order to obtain a visually appealing impression of a digitized version of the object, highly accurate solutions for both 3D shape acquisition and the acquisition of the optical surface reflectance behavior represent crucial prerequisites. However, the complexity of material appearance that is characterized by the complex interplay of surface material, surface geometry and illumination conditions has to be taken into account. While a huge amount of techniques has been developed, the acquisition of surfaces which exhibit a complex surface reflectance behavior is still challenging.

So far, none of the image-based techniques presented in the literature is powerful enough to handle objects made of arbitrary materials. Instead, today's 3D acquisition techniques and reflectance acquisition techniques strongly rely on basic assumptions regarding the surface reflectance behavior. This means that acquisition techniques are rather designed according to the surface materials of the considered objects and, hence, material-specific. After a review of preliminaries of material appearance that include the effects of light exchange at surfaces as well as the dependency of material appearance on material properties, surface geometry and illumination conditions as well as scale, we discuss a taxonomy of surface classes with respect to the characteristics of light transport induced by surface reflectance properties which is relevant for 3D geometry acquisition (see Section 2.2). This is followed by an overview of methods dedicated to 3D acquisition for diverse surface materials (see Section 2.3) that are categorized according to the aforementioned taxonomy as well as an overview on different principles for reflectance acquisition (see Section 2.4). Most of these surveys have already been published in the scope of peer-reviewed tutorials [WK15a, WdBKK15, WLGK16, WWRJ16].

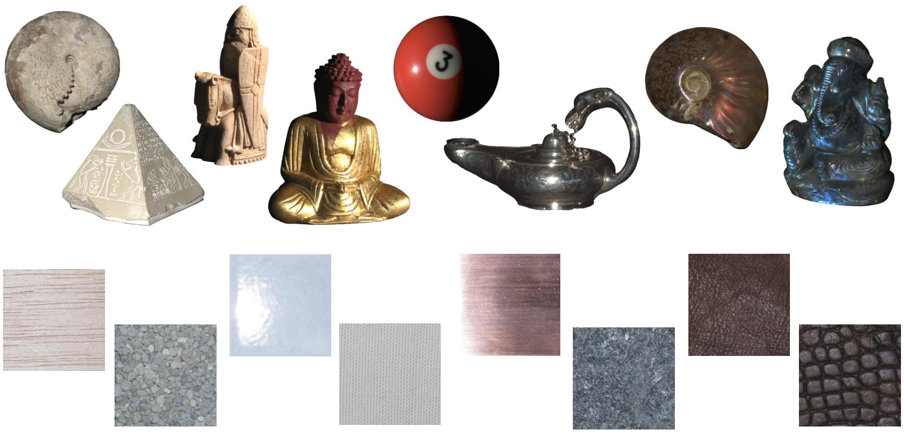


Figure 2.1: *Illustration of several exemplary objects and material samples with different appearance characteristics.*

2.1 Material Appearance

When looking at the materials of objects present in our daily life, we may easily get a first impression of the complexity of visual material appearance. Some exemplary objects are depicted in Figure 2.1. While some of the respective materials are flat, others have a characteristic relief structure. While some of them have only one single color, others are colorful. On some objects, we observe specular highlights or even see reflections of the environment, whereas other objects appear matte. Of course, there are many more such examples in daily life. In order to understand the key effects influencing material appearance, we first might have a closer look at the underlying physical effects that characterize material appearance.

In particular, material appearance is determined by the complex interplay of light, surface geometry and material properties of the object surface. Considering the general case where the incoming radiant flux arrives at the surface at position $\mathbf{x}_i$ at the time t_i with the wavelength λ_i, the flux might enter the material, travel through the material and exit the material at position $\mathbf{x}_r$ at the time t_r with the possibly changed wavelength λ_r. The direction of the incoming flux and the direction of the exiting flux are usually represented based on local coordinate frames that depend on the individual surface points $\mathbf{x}_i$ and $\mathbf{x}_r$ and the local surface normals. Figure 2.2 illustrates this process. Depending on the material type, this rather general 12-dimensional model might be significantly simplified. Typical assumptions made in the great majority of publications are that the light transport at the surface happens in an infinitesimal period (i.e. $t_i = t_r$), that there is no time dependency of the reflectance behavior (i.e. $t_0 = t_i = t_r$), that the wavelength remains unchanged (i.e. $\lambda_i = \lambda_r$) and that the incoming flux is completely reflected at the surface (i.e. $\mathbf{x}_i = \mathbf{x}_r$).

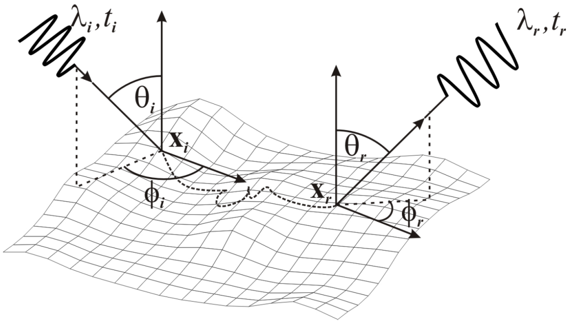


Figure 2.2: *Light exchange at the material surface: The incoming radiant flux hits the material surface at position $\mathbf{x}_i$ at the time t_i with the wavelength λ_i, travels through the material and exits the material at position $\mathbf{x}_r$ at the time t_r with the wavelength λ_r. The incoming direction (θ_i, ϕ_i) and the outgoing direction (θ_r, ϕ_r) of the flux can be formulated using local coordinate frames. Image taken from [MMS$^+$04].*

In this context, it is also necessary to take into account that material appearance is a scale-dependent phenomenon. On a *microscopic scale*, i.e. the scale of atoms and molecules, the interactions of photons with the atoms or molecules of a particular material have been analyzed in the domain of quantum optics. Clearly, these structures cannot directly be observed by the human visual system and yet they significantly contribute to material appearance. While they determine the appearance of all materials, these structures particularly dominate the reflectance behavior of rather smooth and homogeneous materials such as metals, paper, plastics, etc.. On a slightly coarser scale, studies in the field of wave optics have considered the interactions of light with small structures with a size of approximately the wavelength of light to describe effects such as polarization or diffraction. Furthermore, material appearance is also characterized by effects of light exchange happening on a *mesoscopic scale* at fine details in surface geometry such as scratches, engravings, weave-patterns of textiles or embossing of leathers. Such surface structures cause effects like interreflections or self-shadowing. While the effects on these aforementioned scales obviously represent the material characteristics and determine the material appearance, the 3D geometry of the object with the respective, considered material also influences the material appearance significantly. Considering this *macroscopic scale*, regular structures such as present in e.g. woven cloth, brushed metal or surface textures of certain objects might appear distorted in the image because of the dependency on the object geometry.

Unfortunately, the consideration of these scales suffices only for a close distance between the surface material and the human observer. For an increasing distance, the effects of light exchange at fine surface details such as scratches, engravings, weave-patterns or embossing will become less visible and finally, they will not be perceivable as mesostructures anymore. Hence, they might be treated as irregularities in a different kind of microscopic scale. In a similar way, some of the details in the surface geometry might not be perceived as macroscopic features anymore but rather as features on a novel mesoscopic scale.

To give another example, shininess of specular objects or translucency might also depend on the distance between object and observer. When considering a highly specular surface with a rough surface profile from a close range, the resolution of the human visual system is sufficient to perceive the many surface patches with different surface normals, and the material will appear specular. With an increasing distance to the surface, the resolution of the visual system will become insufficient to perceive the appearance of all the individual surface patches with different orientations separately and, instead, a superposition of the appearances of several of these patches is perceived. This will lead to a transition from specular to diffuse appearance perception. In contrast, for flat, highly specular surfaces, the surface will still appear highly specular with an increasing distance due to the rather small deviations of the local surface normals. In a similar way, the appearance of translucent objects with a rough surface profile is characterized by subsurface scattering effects when viewed from a close range and might be perceived as an opaque surface for an increasing distance. For larger distances, only the superposition of the appearances of the individual patches with the subsurface scattering effects is perceived by the visual system.

This clearly indicates that the definition of scale is of dynamic nature. Therefore, material appearance involves a multitude of scales $\ldots \subset D_{i-1} \subset D_i \subset D_{i+1} \subset \ldots$ ranging from an atomic scale to the intergalactic scale [Kaj85, MMS+04].

2.2 A Taxonomy of Surface Classes Based on Light Transport Characteristics

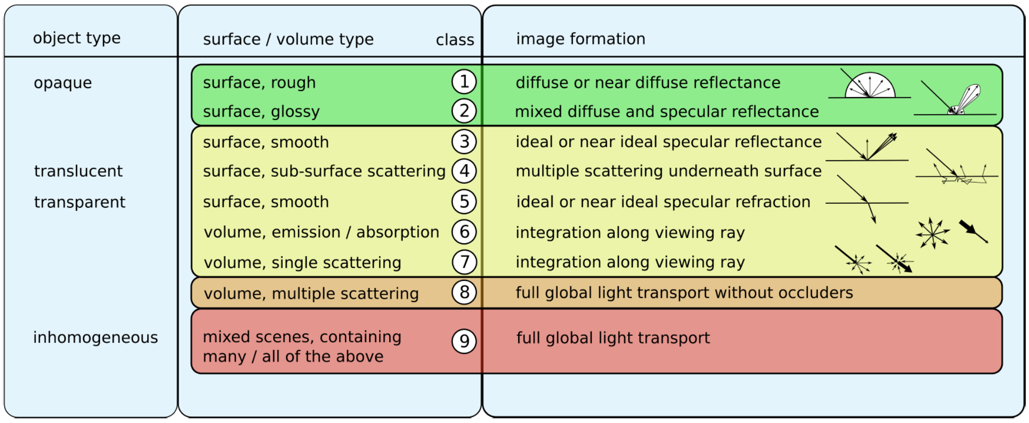
While contact-based 3D geometry acquisition techniques are capable of reconstructing the 3D geometry of almost all solid objects by e.g. using feelers attached to manipulable arms, such approaches are typically considered as impractical. The pointwise surface acquisition results in long acquisition times as dense measurements of the object surface are desirable to accurately reconstruct the surface details. Furthermore, several objects made of e.g. fabrics or clay cannot be acquired contact-based as the pressure of contact might deform or even break them. In particular, the invaluable and typically sensitive cultural heritage artifacts have to be handled as careful as possible which is contradictory to a contact-based acquisition.

As, to the best of our knowledge, none of the available non-contact based acquisition techniques is capable of handling arbitrary materials, the idea of grouping the individual materials according to the acquisition principles applicable to the individual material groups becomes immediately evident. It is easy to imagine that these groups strongly rely on a clustering of the materials with respect to the complexity of their visual appearance. In this spirit, the authors of [IKL$^+$10] consider a taxonomy of object classes based on increasing complexity in light transport as illustrated in Figure 2.3. In particular, nine classes have been identified for categorizing the properties of surface reflectance behavior:

- rough surfaces with diffuse or near diffuse reflectance (see Figure 2.4a)
- glossy surfaces with mixed diffuse and specular reflectance (see Figure 2.4b)
- smooth surfaces with ideal or near ideal specular reflectance (see Figure 2.4c)
- surfaces where light is scattered multiple times underneath the surface (see Figure 2.4d)
- smooth surfaces with ideal or near ideal specular refraction (see Figure 2.4e)
- volumes with light emission or absorption
- volumes where a light ray is scattered only a single time
- volumes where a light ray is scattered multiple times
- mixed scenes containing several of the other types

2.3 The Diversity of 3D Shape Acquisition Methods

The complexity of surface reflectance behavior makes the acquisition of 3D shapes a challenging task and has led to the development of a huge range of diverse material-specific acquisition techniques, each tailored to only a very limited range of materials. Most of the developed methods follow the categorization according to the classes considered in [IKL$^+$10] which have already been discussed in Section 2.2. As we focus on analyzing solid objects in the scope of this thesis, we do not further discuss acquisition techniques designed to handle volumetric phenomena such as fog or fire and only discuss related work in the remaining classes in the following subsections. Furthermore, a complete overview on the large multitude



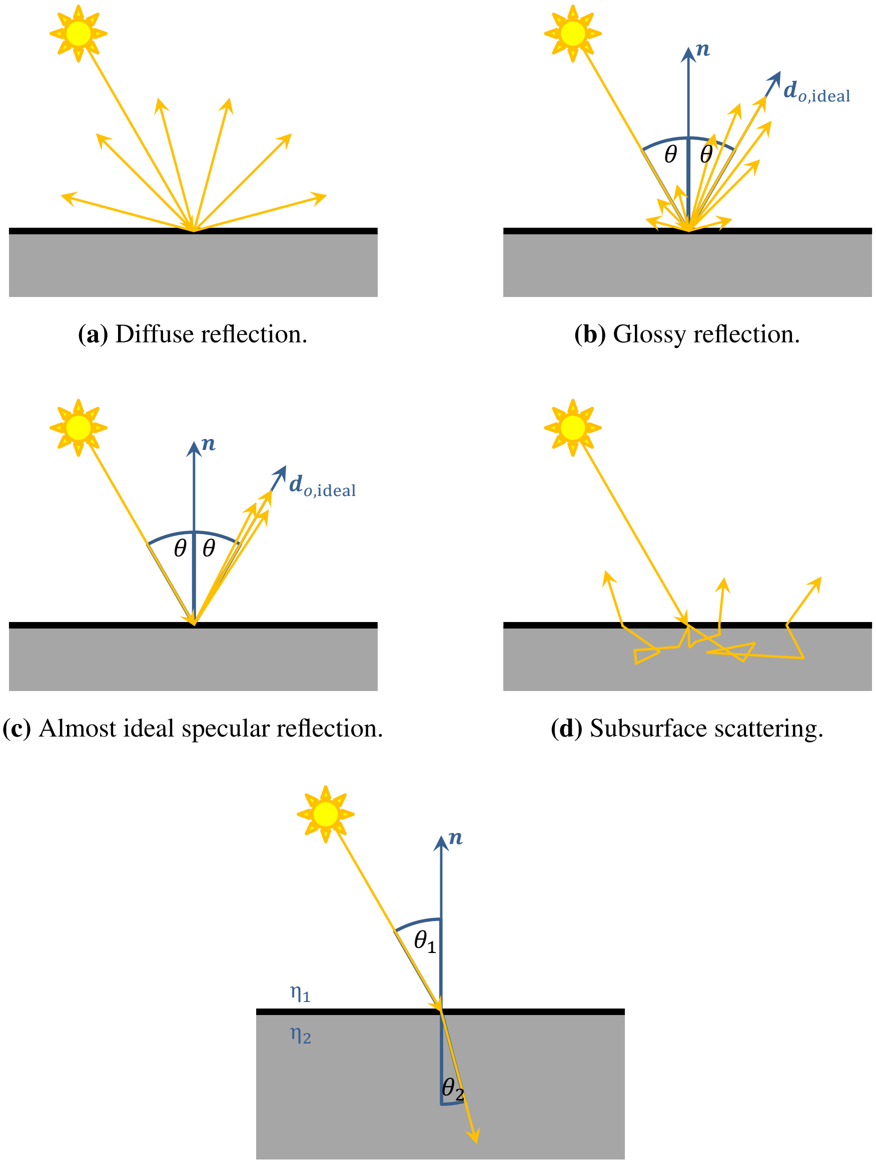
object type	surface / volume type	class	image formation
opaque	surface, rough	1	diffuse or near diffuse reflectance
	surface, glossy	2	mixed diffuse and specular reflectance
	surface, smooth	3	ideal or near ideal specular reflectance
translucent	surface, sub-surface scattering	4	multiple scattering underneath surface
transparent	surface, smooth	5	ideal or near ideal specular refraction
	volume, emission / absorption	6	integration along viewing ray
	volume, single scattering	7	integration along viewing ray
	volume, multiple scattering	8	full global light transport without occluders
inhomogeneous	mixed scenes, containing many / all of the above	9	full global light transport

Figure 2.3: *A taxonomy of object classes according to [IKL$^+$10] (image taken from [IKL$^+$10]).*

of all the developed methods would exceed the scope of this thesis by far. For this reason, we rather give a brief survey on the main trends in shape acquisition according to the survey in [IKL$^+$10] with extensions to more recently published techniques that improved the state-of-the-art.

In general, geometry acquisition methods can be categorized into *passive methods* and *active methods*. Whereas passive methods do not actively manipulate the observed scene, active methods rely on directly manipulating the lighting within an observed scene, e.g. by laser scanning or projecting light patterns.

Some techniques cannot directly be categorized according to this taxonomy as the required information might be obtained in different ways for objects made of different materials. Among these techniques are techniques that use silhouette information to reconstruct the surface geometry. In particular, shape-from-silhouette approaches as used in e.g. [Lau94, GGSC96, FKIS02, MBK05] rely on obtaining accurate silhouette information by a prior segmentation of the respective images into foreground regions that contain the object and background regions that include the remaining scene content. Subsequently, the observed silhouette information is projected into the volume of interest. This can be achieved by using a volumetric, voxel-based representation where the voxels that are projected into the foreground regions within the images of all individual views are considered as occupied and the remaining voxels are set to empty. As a result, the respective 3D surface is represented by the isosurface between the occupied voxels and the empty voxels. However, reliably extracting silhouettes often represents a rather challenging task due to effects such as shadows and occlusions, and, additionally, the properties of the surface materials have to be considered. Moreover, only the shape of objects with a rather simple surface geometry can be accurately acquired. Therefore, this



(a) Diffuse reflection.

(b) Glossy reflection.

(c) Almost ideal specular reflection.

(d) Subsurface scattering.

(e) Specular refraction.

Figure 2.4: *Illustration of different types of surface reflectance behavior for incoming light.*

kind of technique is rather limited and typically has to be combined with additional normal information as in e.g. [CLL07, Dai09] or multi-view stereo consistency as in e.g. [ES04, CK11] to allow the acquisition of concavities in the surface geometry of the considered objects.

2.3.1 Techniques for Rough Surfaces With Diffuse or Near Diffuse Reflectance

For diffuse surfaces (see Figure 2.4a), the incoming light is uniformly reflected into the full hemisphere with respect to the local surface normal of the object geometry. This means that the surface can be perceived in a view-independent way. Therefore, the geometry of such diffuse objects can typically be acquired in a rather easy way and a multitude of acquisition techniques for such surfaces have been developed so far.

Most of the developed geometry acquisition approaches inherently focus on the acquisition of diffuse or near diffuse surfaces by a-priori assuming Lambertian surface reflectance. Amongst others, stereo and multi-view stereo techniques, for which comparisons and performance evaluations can be found in [SS02] and [SCD$^+$06], belong to this group. Multi-view stereo approaches, for which a survey is given in e.g. [SCD$^+$06], allow for the reconstruction of absolute depth. However, these methods rely on the availability of certain characteristic image features which can be reliably detected and used for the establishment of point correspondences across images acquired under different viewing conditions. Based on these point correspondences, the respective point on the object surface can be estimated by triangulating the respective rays passing through the image locations of the individual correspondences and the locations of the projective centers associated with the individual cameras. For this purpose, interest operators such as Harris features [HS88], SIFT features [Low04] and SURF features [BETvG08] represent standard methods for detecting and matching common image contents. However, since these algorithms analyze the local statistics of intensity gradients for the extraction of distinctive feature points and the computation of local descriptors, they can only be applied on textured surfaces and problems arise in the case of an acquisition of objects with homogeneous surfaces, where no particular point “sticks out” of its neighborhood. As a consequence, a robust matching of features often is very difficult. Additionally, these methods are not suitable to capture objects with highly complex reflectance behavior including effects such as specular highlights, transparency, translucency and interreflections as they are – by design – tailored to surfaces exhibiting Lambertian reflectance behavior. To be more specific, multi-view stereo methods are typically based on the assumption that the appearance of a certain point on the object surface does not change under variations of viewing conditions, i.e. the emitted radiance is independent from the view direction. However, this assumption is only valid for diffuse surfaces.

In contrast, active methods are based on generating correspondences by an active illumination of the object surface, which typically leads to a more reliable detection of distinctive features with respect to the matching process. Overviews of current state-of-the-art active geometry acquisition methods, including time-of-flight systems, laser scanners and structured light techniques, are provided in e.g. [Bla04] and [STD09]. While laser scanners rely on successively scanning points on the object surface, structured light systems are based on the idea of projecting known 2D light patterns onto an object surface by projectors. Then, the corresponding patterns that are reflected at the object surface can be observed in the images acquired by cameras. The patterns are used for an encoding of points on the object surface in a way that allows a robust detection of correspondences between points in the image planes of the projectors and points in the image planes of the cameras or between points in image planes of several different cameras respectively. Finally, the object surface can be reconstructed via triangulation. There is a large variety of such approaches regarding the utilized coding strategies that are suitable for different scenarios. A respective survey can be found in [SPB04]. While such active triangulation-based laser scanner systems or structured light systems reach a remarkable scanning accuracy for diffuse surfaces, they can also be used for surfaces with an additional specular component as long as the material still shows a sufficient surface albedo. However, most of these techniques are not robust with respect to effects such as interreflections, subsurface scattering or highly specular surface reflectance. Multi-view observations might provide a certain robustness with respect to these effects. However, if the surface is almost ideally mirroring and does not show a sufficient albedo, the observed pattern information becomes completely view-dependent and cannot be used to establish correspondences across images taken from different views. In a similar way, the correspondences cannot be established in case of transparent or translucent surfaces.

Furthermore, in order to handle objects with non-uniform surface albedos, two different exposure times are used in [SS03] and the one which leads to the largest absolute difference between the two illuminations is selected. For the same reason, the idea of projecting multiple Gray code patterns at different illumination intensities and forming high dynamic scale radiance maps in order to decrease the susceptibility to misclassification caused by the reflectance properties of the considered surfaces has been introduced in [SL00]. Though these approaches address the problem of over- and underexposure, they do not compensate for the non-linearity of the response curve of the camera. In contrast, in our approach described in Chapter 3, we propose to overcome this problem by using HDR images.

An alternative approach has been proposed in [KPDVG05] where the projector intensities are adapted locally so that the dynamic range of the illuminated scene is reduced in order to avoid over- or underexposure. However, this method suffers from a decreased contrast in the adapted projector pixels. This problem does not occur when, instead of adapting the dynamic range of the scene, the scene is completely captured by taking a sequence of images with different exposure times. In addition, it is desirable to further increase the resolution of structured light based techniques to further improve their accuracy. In e.g. [RSGS10], lens-shifting has been exploited to enhance the resolution. Similar to [ALY08], we consider increasing the resolution in the context of a multi-projector, multi-camera setup. While projectors have been used to actively illuminate the object of interest or as virtual cameras in [ALY08], our super-resolution technique presented in Chapter 3 and published in [WSRK11] exploits overlaps of the pattern projections for projectors placed at different locations.

Amongst the early investigations regarding the use of normal information for 3D reconstruction of diffuse objects are shape-from-shading techniques which can be dated back to the investigations in [Hor70]. These passive approaches are based on an analysis of the shading gradient in an image that can be used to derive surface normals. Based on the surface normals, the surface geometry is derived via normal field integration techniques. Most of the shape-from-shading techniques focus on geometry reconstruction from a single-image, and, hence, only a 2.5D height map can be derived. Further early investigations include photometric stereo techniques [Woo80] which focus on reconstructing Lambertian objects from a single view under known positions of the light source involved in order to illuminate the object. Several techniques such as the ones proposed in e.g. [BJK07, WGS$^+$11, WLDW11] consider extending photometric stereo towards general unknown illumination. Furthermore, photometric stereo has been explored in multi-view setups (e.g. [EVC08, BAG12]). The approach presented in [BAG12] is based on using normal consistency by determining a maximal set of inliers per voxel on which regular photometric stereo is applied in a multi-view approach. While producing good reconstructions on synthetic data, the estimated surface consistency tends to being localized inaccurately for real-world data due to the lack of a per-voxel normalization. Furthermore, multi-view normal field integration approaches as proposed in [CLL07, Dai09] have been considered in the context of photometric stereo. These techniques overcome the problem of obtaining only 2.5D reconstructions of partial surfaces in the single-view case. In [CLL07], an initial visual hull reconstruction is followed by an iterative surface evolution based on level sets in a variational formulation. As no global optimization is performed, the surface evolution is sensitive to the initial visual hull. In contrast, the technique proposed in [Dai09] is based on a Markov Random Field (MRF) energy function where the surface is computed via min-cut to find a global minimum. This is followed by a smoothing step similar to the one applied in [CLL07]. A surface orientation constraint has been included in the energy functional which enforces the reconstructed geometry to agree with the observed surface normals. Both techniques employ additional silhouette information which are rather difficult to determine.

In Chapter 4, we will demonstrate that combining a structured light technique with additional normal information even allows the reconstruction of very fine surface details such as scratches or engravings on both synthetic and real-world objects.

2.3.2 Techniques for Glossy Surfaces With Mixed Diffuse and Specular Reflectance

Considering materials with both diffuse and specular components (see Figure 2.4b) makes the 3D shape acquisition more complex as the perception of glossy highlights is a view-dependent phenomenon. With the objective of also acquiring such objects with more complex surface reflectance behavior, several methods considered extending classical photometric stereo. Therefore, violations of the underlying assumption of Lambertian reflectance due to specularities have to be considered as well as shadows. In e.g. [HS05, GCHS05], spatially-varying BRDFs have been considered to broaden the range of materials that can be handled, but several effects such as e.g. interreflections and shadows are not taken into account. The extension of photometric stereo proposed in [CJ08] is based on a normal estimation using shadow boundaries and the estimation of a BRDF where the Ward model [War92] is used. The obtained BRDF parameters are then clustered to find the material types and discard noise and outlier values. Once the specular parameters are known, their method proceeds with estimating normals and surface albedo. Limitations of this method can be seen in the use of a Ward BRDF model, which is only valid for a rather small range of materials. Therefore, this technique is only capable of handling isotropic materials with a single lobe. Materials for which more than a single lobe has to be fitted as well as the huge range of materials which exhibit mesoscopic effects cannot be handled appropriately. Furthermore, only a 2.5D reconstruction from a single view is performed based on an orthographic projection model and normal field integration techniques are susceptible to errors in the estimated normals that are accumulated during the surface reconstruction. In [ZMLC10], photometric stereo has been extended by representing specular reflection with a set of specular basis functions with different roughness values. Therefore, specular objects can be handled as well by their photometric stereo technique and the final reconstruction consists of information regarding surface shape and also reflectance properties. However, the range of materials that can be handled is limited to isotropic materials that can be represented by the Ward model with the specular basis functions. Furthermore, only a 2.5D reconstruction from a single view is considered and an orthographic camera projection is assumed. In [RK09], a combination of multi-view stereo and photometric stereo has been used to directly reconstruct heightfields without the need to apply normal integration, and a SVBRDF is recovered as well. While this approach also takes interreflections into account, it is limited to planar samples and not applicable to objects with a complex surface geometry.

In order to handle the even larger range of opaque materials, the approaches presented in [ZBK02, ZHK$^+$03] exploit the Helmholtz reciprocity for surface normal estimation in a BRDF-invariant manner. This principle is based on the observation that for certain materials the ratio of the radiance emitted from a surface point towards the camera to the irradiance that arrives at the surface point from the direction of the light source is identical if the positions of camera and light source are exchanged [ZBK02]. In [DPB10], Helmholtz normals have been used in a multi-view setting.

Several other approaches focus on improving photometric stereo by considering normal hypotheses in a volumetric representation and the surface is assumed to pass through voxels with a high consistency of the hypothesized normals. In e.g. [CJ82, MC09], classical single-view photometric stereo has been improved by selecting only hypotheses which agree with the underlying model assumptions. Generating per-pixel normal hypotheses for varying lighting directions has also been used in [HMI10] where consistency is obtained by considering monotonicity, visibility, and isotropy properties. Therefore, the approach can handle both diffuse and specular surfaces.

Furthermore, structured light techniques, as already mentioned in Section 2.3.1, are also applicable for materials with both a diffuse and a specular reflectance component as long as the surface material shows a sufficient albedo under different viewing conditions. In these cases, the projected patterns can still be reliably decoded in the acquired images. Some techniques focus on further increasing the robustness with respect to the range of materials that can be coped with by using pattern projections at different intensity levels [SL00] or by capturing images for different exposure times [SS03]. In particular, in Chapter 3, we will demonstrate that even the shape of moderately specular objects can be accurately acquired with our structured light system described in Chapter 3 and [WSRK11] – which is also based on high dynamic range imaging – as long as the material still shows a sufficient surface albedo. Limitations of structured light approaches as mentioned in the previous section are the lacking robustness with respect to effects such as interreflections, subsurface scattering, or highly specular surface reflectance. To some degree, using multi-view observations might improve the robustness with respect to these effects. However, for almost perfectly specular surfaces with a low albedo, the pattern information observed by the cameras is view-dependent. This makes the establishment of correspondences across images obtained under different view conditions based on the observed pattern information impossible.

In addition, we present a technique for even more accurate geometry acquisition for the full range of opaque objects from diffuse to glossy objects in Chapter 4 and [WRO+12]. Our method is based on positional information obtained with a structured light technique and normal information derived based on Helmholtz reciprocity. This overcomes the noise introduced in purely triangulation-based approaches such as structured light systems due to inaccurately localized correspondences and inaccuracies in calibration, which would smooth away fine surface details such as scratches or engravings.

2.3.3 Techniques for Smooth Surfaces With Ideal or Near Ideal Specular Reflectance

The challenge in reconstructing highly specular surfaces arises from the fact that such surfaces typically do not have an own characteristic appearance but rather reflect the surrounding environment in a view-dependent manner (see Figure 2.4c). Recent surveys on techniques for surface reconstruction of specular objects have been presented in [IKL+10, BW10]. Furthermore, a more theoretical discussion is given in [KS08].

Some of the methods, such as specular flow techniques [RB06, AVBSZ07], are based on considering the movement of environment features which are mirrored on the surface of specular objects. Typically, a known motion of the mirroring object, its environment or the cameras is assumed respectively. Unfortunately, interreflections might cause a single environment feature to be observed on the mirroring surface several times which makes the estimation of dense optical flow highly non-trivial. In addition, such methods usually rely on the assumption of a distant environment and do not consider a more complex scene geometry. Instead of considering dense correspondences, the approach in [SVTA10] is based on using sparse reflectance correspondences to locally approximate specular surfaces using quadrics.

Furthermore, shape acquisition of mirroring objects has been approached by utilizing the information revealed by specular highlights which are observed on the mirroring object surface due to specular reflection in controlled environments. However, densely sampled observations of specular highlights on the mirroring surface are required to obtain an accurate reconstruction of the surface geometry. This can be achieved by using a moving camera [ZGB89], moving the light source [CGS06], using extended light sources [Ike81] or sequentially switching on individual elements of a grid of light sources [SWN88].

As the number of required images increases linearly with the number of utilized light source positions, taking one photo for each of the utilized light source positions would be impractical for the dense sampling of the light source positions required to obtain an accurate reconstruction. For this reason, several techniques focus on the reduction of the number of required images by performing measurements in parallel. For this purpose, rotating the object and using a circular light source has been proposed in [ZM00]. Furthermore, printed, static or moving calibrated patterns have been used in [BS03, SCP05, LWDC10]. Further methods make use of the simultaneous encoding of multiple light sources. Such encoding schemes have already been investigated in [NWSS90] for light source arrays, and several more recently published approaches build upon this idea by simulating dense illumination arrays using LCD screens and encoding the illumination emitted from the pixels using structured light patterns [FCM$^+$08, NWR08, YIX07, BHB11]. So far, most of the approaches still rely on assuming far-field illumination or a distant environment. However, these assumptions are typically not fulfilled as the printed patterns or the LCD displays used to project patterns have to be located rather close to the object surface in order to obtain dense observations of light directions or feature directions on the mirroring surface.

Knowing the view direction that is determined by the camera parameters and the pixel-based observation of a specular highlight, and the 3D position of the light source or feature on the utilized patterns is still not sufficient to infer the surface geometry due to a remaining normal-depth ambiguity. This ambiguity can be overcome in a multi-view setting as e.g. presented in [BS03], where a calibrated pattern is used to produce reflections on the specular surface. Based on a volumetric representation, the law of reflection is used to hypothesize a normal at each voxel. As a result, generally several normal hypotheses are obtained for the different view directions and light source positions or feature positions. This allows to approach surface reconstruction by assuming that the surface passes through the voxels with the most consistent normal hypotheses. The respective normal consistency is computed per voxel according to a normal disparity measure. As voxels with a low consistency are discarded, such principles are often named as *voxel carving techniques*. The technique proposed in [Pak12] is also based on voxel carving but considers the consistency of normal vector maps per voxel that are used to parameterize the normal hypotheses per voxel. Several other works such as e.g. [CJ82, MC09, HMI10, BAG12] also consider techniques based on hypothesizing surface normals in the context of extending photometric stereo techniques (see

Section 2.3.2). For specular surfaces, investigations on surface reconstruction based on the idea of matching hypothesized normals include the approaches proposed in [WI93, BHB11]. While considering overlapping deflectometric measurements obtained from multiple views can be used to reconstruct large mirroring surfaces as shown in [BHB11], self-occlusions are problematic for this approach and a lot of manual work is involved in configuring the individual views.

The technique presented in [YWT+11] considers normal consistency in a single-view setting by clustering per-pixel normal observations using the k-means algorithm [Ste57, Llo57, Llo82, Mac67]. In [NWR08], a specularity consistency similar to the one in [BS03] is derived between a set of views in a triangulation-based scheme using a display with Gray codes for illumination. After triangulation, normals are refined for the estimated depth values in a way similar to the iterative scheme proposed in [TLGS05]. In [Pak13], a probabilistic voxel carving technique has been presented that uses an optimization based on loopy belief propagation. However, only synthetic data has been considered and, even in this rather ideal scenario, the reconstruction results are rather inaccurate.

Furthermore, the technique recently published in [TFG+13] focuses on the acquisition of objects with diffuse to specular surface reflectance behavior by using continuous spherical harmonic illumination. The response of the object to the harmonics can be used to separate the diffuse reflectance component from the specular reflectance component. Unfortunately, this technique is less suitable for the reconstruction of concave objects.

Our method presented in Chapter 5 and also published in [WORK13], to the best of our knowledge, represents the state-of-the-art technique as it allows an accurate acquisition of the full 3D shape of mirroring objects with complex surface geometry as present in many figurines or artifacts. In particular, the involved multi-view normal field integration scheme is probably the first of its kind which enables an accurate 3D reconstruction not only on synthetic data but also on real-world data.

2.3.4 Techniques for Surfaces With Subsurface Scattering Effects

The challenge in acquiring the 3D shape of translucent objects arises from the light transport within the object (see Figure 2.4d). In particular, the incoming light enters the material and travels through the material where it is scattered. When we actively illuminate such translucent objects with a pattern, these non-local subsurface scattering effects induce a blurring of the observed pattern and, hence, make e.g. a triangulation-based reconstruction from the decoded correspondences rather unreliable [LPC+00].

In [GLL+04], the authors circumvent these problems arising from the subsurface scattering characteristics by covering the object of interest with a thin, diffuse dust before the actual 3D geometry acquisition via laser scanning is started. Later, this dust can easily be removed again.

Apart from using such tricks, surface geometry of translucent objects can also be acquired by utilizing certain material-specific characteristics of light transport. As many translucent objects also have a strong specular component, shape-from-specularity approaches can be applied where a moving light source is involved and the observed highlights can be used to estimate surface normals which is followed by a normal field integration [CGS06]. In [MHP+07], linearly polarized and circularly polarized spherical gradient illumination patterns are used and both the diffuse and the specular reflectance is considered for the estimation of surface normals. The advantage of the circularly spherical patterns can be identified in the fact that they allow the simultaneous estimation of surface normals from different viewpoints. The proposed polarized illumination schemes allow an independent estimation for both diffuse and specular normal maps. The latter have been proven to be appropriate for subsurface scattering materials in contrast to the diffuse normal estimates which are affected by the subsurface scattering. More recently, this method has also been used in [GCP+10] with circularly polarized spherical illumination for normal estimation from several viewpoints.

Furthermore, the investigations in [NKGR06] have demonstrated that specular and diffuse components of surface reflection can be separated by phase shifting of high-frequency structured light patterns. This observation has e.g. been explored in [CLFS07], where a phase shifting based structured light approach has been combined with a polarization-based removal of specular highlights at the surface. Based on the fact that global light transport characteristics remove the polarization of light, polarization filters are used in front of both the light source and the camera and multiple scattering effects can be separated from the structured light observations. In subsequent work [CSL08], the same authors remove the dependency on polarization and instead modulate the low-frequency phase shifting patterns to separate direct and global components of light transport. By this modification, the obtained reconstruction quality is further improved in comparison to the technique in [CLFS07].

In [GAVN11], certain structured light patterns tailored to translucent surfaces have been proposed. While the analysis shows that high-frequency patterns are not applicable for translucent objects due to the blurring of the observed pattern, using Gray codes with a certain minimum stripe width following [GG03] shows more robustness on translucent surfaces and allows reliable reconstructions.

The recent technique presented in [DMZP14] represents an extension to conventional photometric stereo which enables the simultaneous estimation of both scattering properties and accurate surface normals for planar, homogeneous translucent objects based on observations from at least three different directional illumination configurations based on blind deconvolution. This avoids the problem of blurry normal estimates that would result from an acquisition via conventional photometric stereo.

2.3.5 Techniques for Smooth Surfaces With Ideal or Near Ideal Specular Refraction

Reconstructing the 3D shape of refractive objects (see Figure 2.4e) is even more challenging in comparison to the cases mentioned in the previous sections. In general, such objects might exhibit inhomogeneous reflectance characteristics induced e.g. by a spatially varying refractive index or by inclusions of Lambertian or opaque material components. As pointed out in the recent survey given in [IKL$^+$10], research has mainly been spent on solutions relying on certain simplifying assumptions such as homogeneous material characteristics or considering only the reconstruction of a single surface separating the two enclosing media. The authors categorize the main approaches for the acquisition of refractive surfaces according to shape-from-distortion techniques, direct ray measurement techniques, reflectance-based techniques, techniques based on inverse ray tracing, tomography-based approaches and direct sampling techniques. In the scope of this section, we therefore group the related approaches according to these principles.

For the simpler case of acquiring a single refractive surface, shape-from-distortion techniques have been successfully applied. While this kind of methods can also be applied for specular surface reconstruction in a simpler form, refractive surface reconstruction requires considering the refractive index in addition to the surface normal in order to analyze the light path. Early work [Mur90, Mur92] has focused on reconstructing water surfaces from a single view. The movement of the water induces the a-priori unknown background pattern placed at the bottom of the liquid to be observed in a distorted way. Assuming an orthographic camera, optical flow [HS81, LK81] and a subsequent integration of the surface gradient are used to reconstruct the surface up to a certain scale. This seminal work has been extended in [MK05] by using a stereo camera system and a known pattern to estimate the refractive index, per-pixel depth and surface normals. As a further improvement, no average surface model is used in comparison to the approaches in [Mur90, Mur92]. In particular, the considered consistency measure is similar to the normal consistency used in [NWR08] for specular surface reconstruction. Further work has been dedicated to the reconstruction of glass objects. Projecting structured light patterns into the refractive object with a projector and observing the respective distorted patterns in the camera image has been analyzed in [HSKK96]. In [BEN03], an unknown distant background pattern is used in combination with a known parametric model including shape and refractive index. The object of interest is moved in front of a single, static camera and features are tracked over time similar to [Mur90, Mur92]. In [AMKB04], an extension of optical flow has been proposed to track refracted scene features for which the intensity might vary due to the presence of non-ideally transparent surfaces with e.g. additional absorption.

Furthermore, refractive surfaces have also been reconstructed by directly measuring the light path. For this purpose, calibrated planar patterns in several positions with respect to the object have been used in [KS05, KS08] to measure the light rays. In their theoretical analysis [KS05, KS08], the authors consider the categorization of reconstruction techniques based on ray measurements independently performed for each pixel. The introduced notation $\langle N, M, K\rangle$ contains the relevant information with respect to the number of views N that are required for the reconstruction as well as the number M of points on specular or refractive surfaces that are located on a piecewise linear light path and the number K of calibrated reference points on a ray exitant from the object. The authors discuss that such a reconstruction cannot be performed for more than two intersections of the light ray with specular or refractive surfaces. The number of views N and the number K of calibrated reference points on a ray do not influence this observation. Following this concept, the authors consider a $\langle 3, 2, 2\rangle$ reconstruction for refractive surfaces. As a result, four surface points with attached normals can be estimated per pixel. While one such pair of point and normal is located at the front surface of the object, the remaining pairs depend on the three differently refracted viewing rays and are located at the back surface.

A reflectance-based approach has been followed in [MK07]. Dense per-pixel reflectance measurements in a static camera are observed as a result of sequentially illuminating the static object of interest with a light source at varying positions on a regular grid. As a result, a 2D slice of the BRDF is recorded. However, indirect lighting effects influence these measurements. By separating the direct and indirect components of light transport, the authors achieve high-quality reconstructions for depth and normal, even for inhomogeneous refractive objects.

Other methods rely on the principle of inverse ray tracing. The underlying idea is based on the optimization of the residual of the acquired data and synthetically generated data. In order to reconstruct the surface of time-varying water surfaces, the water has been mixed with a fluorescent dye in [IM05] and a chemiluminescent chemical in [GILM07]. While using UV illumination makes the mixture of water and fluorescent dye self-emissive [IM05], in the case of chemiluminescence [GILM07] a chemical process has to be used for this purpose. Assuming homogeneous emission, both methods use synthetic images for surface fitting via level set optimization. In [WLZ+09], the liquid to be reconstructed is dyed with an opaque white paint. As a result, patterns can be projected onto the liquid and the correspondences observed in the cameras allow the reconstruction of the surface. In addition, a physically-based fluid simulation is used in this approach.

As discussed in [IKL+10], refractive object reconstruction can be performed by making use of certain acquisition strategies. One possibility is to consider sufficiently high wavelengths for the incident illumination as given for x-rays. In this spirit, computer tomography has been used for scanning objects in [KTM+02], and the proposed approach is in principle capable of scanning glass objects. Furthermore, as mentioned in [IKL+10], a reconstruction of refractive objects is also possible when the refractive indices of the object and the surrounding medium are identical. In [TBH06], glass objects are immersed into a liquid with carefully controlled refractive index. Controlling the refractive index to a value of approximately 1.55 has been achieved by adding chemicals to water. As an ideal transparent object would disappear inside a medium with identical refractive index, the surrounding medium has to be dyed, which can be omitted if the object itself is absorptive [IKL+10].

Furthermore, several methods have been proposed based on direct sampling. In [HFI+08], fluorescent immersion range scanning has been proposed to reconstruct refractive objects. The objects are placed in a different immersing medium with known refractive index. This liquid has additionally been dyed with a fluorescent chemical. During measurement, this fluorescent liquid causes the utilized laser sheet to be rendered visible while the refractive object to be scanned remains dark. A similar strategy has been explored by observing objects in a different spectrum. Several investigations e.g. focus on scanning-from-heating. In [EAM+09], the glass surface is first heated by the incident infrared radiation which is then measured by an infrared camera. This allows to reconstruct the glass surface. Extending this approach, the shape of the hot spots observed by the infrared camera is analyzed in [AEB+12] to derive information regarding local surface orientations. Other works consider structured patterns of infrared radiation [MSSE+10] or the polarization in the infrared domain [MRA+12] for 3D reconstruction. Furthermore, several publications focus on exploiting ultra-violet radiation. In e.g. [RSFM10b, RSFM10a], structured light in the ultra-violet domain has been explored for reconstructing glass surfaces.

2.4 Reflectance Acquisition Techniques

While reflectance acquisition is not in the focus of this work, we nevertheless give a short overview on different reflectance acquisition approaches, as the selection of appropriate acquisition techniques presented in Part IV of this thesis also has a significant influence in this regard.

For an adequate acquisition of surface reflectance, the complexity of visual surface reflectance has to be considered in a similar way as in the context of geometry acquisition. The categorization of materials as discussed in Section 2.2 indicates that visual material appearance is characterized by different phenomena of light exchange with a particular object surface of interest, which might also be explored when focusing on reflectance acquisition. In particular, diffuse and specular components as well as potentially occurring subsurface scattering or refraction characteristics have to be considered in the reflectance models, and the respective reflectance acquisition is typically designed according to the assumed underlying model. So far, many different models have been proposed in the literature to model surface reflectance behavior, each focusing on accurately representing a certain subset of the possible materials. However, efficiently modeling surface reflectance behavior is also coupled with the use of an adequate model, which should have as few parameters as possible to still enable a faithful depiction of the material in a synthetic image within an acceptable acquisition time. Therefore, modeling e.g. the surface reflectance behavior of a diffuse object, where the incoming light is reflected uniformly into a hemisphere on the local surface patch, requires considering different material characteristics than modeling surface reflectance of mirrors, which is determined by an almost ideal direct reflection of the incoming light. Similarly, modeling reflectance behavior for materials with both diffuse and specular components or translucent and transparent materials requires considering the respectively relevant characteristics of the individual materials. In this regard, reflectance acquisition strongly depends on the representation used to model the reflectance of a particular material, as some parameters might not have to be measured. For e.g. a diffuse material, there is no need to capture the specular characteristics, which allows to reduce the hardware usage and, hence, to speed up the measurement process.

A rather general way to model surface reflectance can be seen in using a function

$$\rho(\mathbf{x}_i, \theta_i, \varphi_i, t_i, \lambda_i, \mathbf{x}_r, \theta_r, \varphi_r, t_r, \lambda_r), \tag{2.1}$$

which captures reflectance depending on the twelve parameters of the incoming light direction (θ_i, φ_i), the position $\mathbf{x}_i$ where the light hits the surface at the time t_i with the wavelength λ_i, the position $\mathbf{x}_r$ where the light exits the surface with the outgoing light direction (θ_r, φ_r) at the time t_r with the wavelength λ_r (see Section 2.1). Typically, the time and the wavelength characteristics are omitted for simplicity, i.e. it is assumed that $t_i = t_r$ and $\lambda_i = \lambda_r$.

The plenoptic function $P(\mathbf{X}, \theta, \varphi)$ has been introduced in [AB91] as a function that describes the radiance sent from a certain scene point $\mathbf{X} \in \mathbb{R}^3$ into the direction (θ, φ). Assuming that an arbitrarily complex shaped surface S with the bounding volume V is embedded in the considered scene volume, the radiance values observed at points $\mathbf{x} \in \partial V \subset \mathbb{R}^2$ on the surface from a viewpoint outside V with viewing directions (θ_r, φ_r) can be used to represent the appearance of the object surface S for a given, static illumination [GGSC96, LH96]. As the surface points can be parameterized over the surface S, it is possible to use a four-dimensional function $\rho_{\text{SLF},r}(\mathbf{x}, \theta_r, \varphi_r)$, the radiant surface light field, instead of the five-dimensional plenoptic function $P(\mathbf{X}, \theta_r, \varphi_r)$. If the viewpoint is inside of the volume V, the incident surface light field $\rho_{\text{SLF},i}(\mathbf{x}, \theta_i, \varphi_i)$ can be observed [LH96]. Consequently, a radiant light field $\rho_{\text{LF},r}(\mathbf{x}_r, \theta_r, \varphi_r)$ observed at particular scene points $\mathbf{x}_r$ from the viewing angles (θ_r and φ_r) can be interpreted as a general response of the arbitrary complex scene to an incoming light field $\rho_{\text{LF},i}(\mathbf{x}_i, \theta_i, \varphi_i)$ [LH96]. From the definition of surface light fields it becomes obvious that they can only be used to describe static scenes without variations in illumination, scene geometry and surface materials. The term *reflectance field* [DHT$^+$00] describes the dependency of the radiant light field $\rho_{\text{LF},r}(\mathbf{x}_r, \theta_r, \varphi_r)$ on the incident light field $\rho_{\text{LF},i}(\mathbf{x}_i, \theta_i, \varphi_i)$ and can be formulated as an eight-dimensional function

$$\rho_{\text{RF}}(\rho_{\text{LF},i}, \rho_{\text{LF},r}) = \rho_{\text{RF}}(\mathbf{x}_i, \theta_i, \varphi_i, \mathbf{x}_r, \theta_r, \varphi_r), \tag{2.2}$$

i.e. a reflectance field describes the exitant radiance depending on the possibly occurring incident illuminations. Typically, reflectance fields are defined on convex surfaces that surround the respective object and it is assumed that the viewpoint is outside this bounding volume and that the light is also coming from outside. This allows to use the reflectance field to represent material appearance under arbitrary new viewpoints and illumination conditions by sampling the outgoing light fields under a set of basis incident light fields. The linearity of light transport allows the definition of new illumination conditions in terms of a linear combination of the illumination basis. The definition (2.2) is closely related to the definition of the *bidirectional scattering-surface reflectance distribution function (BSSRDF)* [NRH$^+$77] given by

$$\rho_{\text{BSSRDF}}(\mathbf{x}_i, \theta_i, \varphi_i, \mathbf{x}_r, \theta_r, \varphi_r) \tag{2.3}$$

and even would be identical if the true surface is used.

Assuming that the surface reflectance is defined on the object surface similar as for BSSRDFs but assuming $\mathbf{x}_i = \mathbf{x}_r = \mathbf{x}$, i.e. that light is not scattered inside the material, the inhomogeneous reflectance behavior can be explained by six-dimensional *spatially-varying bidirectional reflectance distribution functions (SVBRDFs)*

$$\rho_{\text{SVBRDF}}(\mathbf{x}, \theta_i, \varphi_i, \theta_r, \varphi_r). \tag{2.4}$$

In contrast, another six-dimensional representation is given by *bidirectional texture functions (BTFs)* which can be obtained by assuming far-field illumination where the light sources are infinitely far away. This means that the incident radiance is the same for all surface points, i.e. $\rho_{LF,i}(\mathbf{x}_i, \theta_i, \varphi_i) = \rho_{LF,i}(\theta_i, \varphi_i)$. As a result, the definition of the BTF is given by

$$\rho_{\text{BTF}}(\theta_i, \varphi_i, \mathbf{x}_r, \theta_r, \varphi_r). \tag{2.5}$$

In comparison to SVBRDFs, BTFs allow to capture local subsurface scattering characteristics as well as mesoscopic effects such as interreflections, self-masking or self-occlusions.

In contrast, when assuming homogeneous reflectance behavior, the BSSRDF can be relaxed to the *bidirectional subsurface scattering distribution function (BSSDF)*

$$\rho_{\text{BSSDF}}(\theta_i, \varphi_i, \mathbf{x}_r - \mathbf{x}_i, \theta_r, \varphi_r), \tag{2.6}$$

which still is capable of modeling subsurface-scattering effects. Additionally assuming non-subsurface scattering reflectance, the *bidirectional reflectance distribution function (BRDF)*

$$\rho_{\text{BRDF}}(\theta_i, \varphi_i, \theta_r, \varphi_r) \tag{2.7}$$

models the reflectance behavior under the remaining four parameters. Considering BTFs or SVBRDFs and additionally assuming homogeneous surface reflectance also leads to the BRDF model. Assuming non-anisotropic reflectance, isotropic BRDF models represent a further simplified reflectance model. In addition, the diffuse surface reflectance function represents a further four-dimensional representation depending on the parameters $\theta_i, \varphi_i, \theta_r$ and φ_r to model diffuse surfaces. Further non-material dependent simplifications of BTFs, as illustrated in Figure 2.5, are the restriction to a fixed lighting or a fixed view resulting in four-dimensional *surface light fields (SLFs)*

$$\rho_{\text{SLF}}(\mathbf{x}, \theta_r, \varphi_r) \tag{2.8}$$

or *surface reflectance fields (SRFs)*

$$\rho_{\text{SRF}}(\mathbf{x}, \theta_i, \varphi_i) \tag{2.9}$$

respectively. In the case of diffuse reflectance, this can be further simplified to two-dimensional *texture maps* or *bump maps*

$$\rho_{\text{Texture Map/Bump Map}}(\mathbf{x}). \tag{2.10}$$

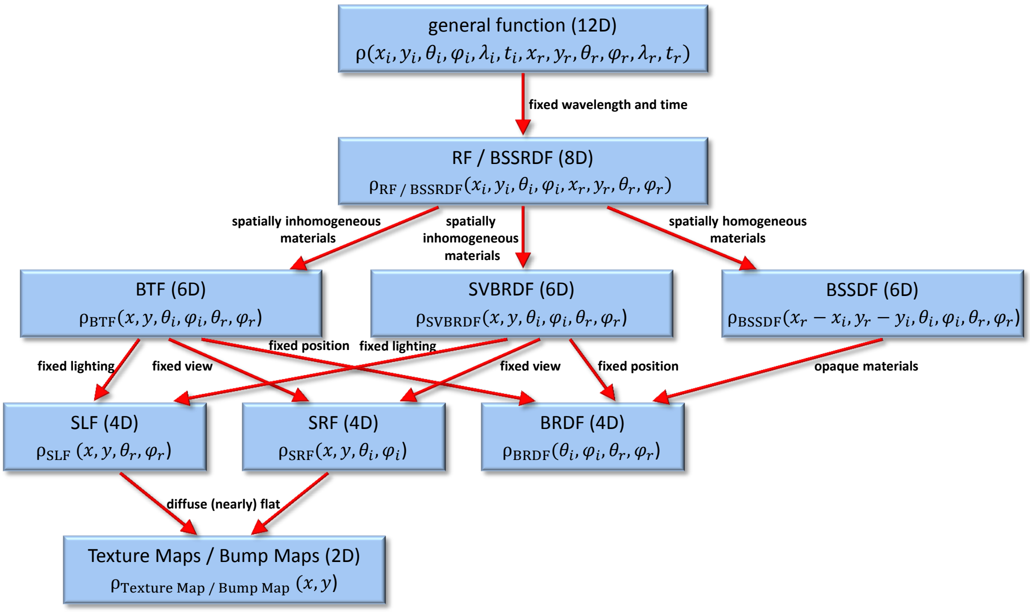

general function (12D)
$\rho(x_i, y_i, \theta_i, \varphi_i, \lambda_i, t_i, x_r, y_r, \theta_r, \varphi_r, \lambda_r, t_r)$
fixed wavelength and time
RF / BSSRDF (8D)
$\rho_{\mathrm{RF\,/\,BSSRDF}}(x_i, y_i, \theta_i, \varphi_i, x_r, y_r, \theta_r, \varphi_r)$
spatially inhomogeneous materials
spatially inhomogeneous materials
spatially homogeneous materials
BTF (6D)
$\rho_{\mathrm{BTF}}(x, y, \theta_i, \varphi_i, \theta_r, \varphi_r)$
SVBRDF (6D)
$\rho_{\mathrm{SVBRDF}}(x, y, \theta_i, \varphi_i, \theta_r, \varphi_r)$
BSSDF (6D)
$\rho_{\mathrm{BSSDF}}(x_r - x_i, y_r - y_i, \theta_i, \varphi_i, \theta_r, \varphi_r)$
fixed lighting
fixed view
fixed position
fixed lighting
fixed view
fixed position
opaque materials
SLF (4D)
$\rho_{\mathrm{SLF}}\ (x, y, \theta_r, \varphi_r)$
SRF (4D)
$\rho_{\mathrm{SRF}}(x, y, \theta_i, \varphi_i)$
BRDF (4D)
$\rho_{\mathrm{BRDF}}(\theta_i, \varphi_i, \theta_r, \varphi_r)$
diffuse (nearly) flat
Texture Maps / Bump Maps (2D)
$\rho_{\mathrm{Texture\ Map\,/\,Bump\ Map}}\ (x, y)$

Figure 2.5: *A hierarchy of reflectance functions according to [MMS+04] and [DLG13].*

This hierarchy of commonly used reflectance functions according to [MMS+04, DLG13] is shown in Figure 2.5. Depending on the parameters of the respective material model, the acquisition device has to be designed in a way that material appearance can be captured under the involved parameter configurations. Further details regarding the individual reflectance models and their acquisition are discussed in [WK15a, WLGK16].

In the following, we will consider a few examples for appropriately representing some materials. When considering a flat piece of specular metal, parametric BRDF models might be a good choice which means that acquisition has to consider material appearance depending on the incoming light direction and the outgoing light direction. When considering a brushed metal, spatial variations in material appearance have to be additionally taken into account which leads to the use of SVBRDFs. In contrast, when considering materials such as leather samples that exhibit spatially varying mesoscopic effects such as interreflections and self-shadowing, parametric models are not sufficient to capture these effects. Instead, it is a better choice to consider data-driven BTF models which are well-suited for the representation of such effects.

CHAPTER 3

A SUPER-RESOLUTION FRAMEWORK FOR STRUCTURED LIGHT

In the scope of this chapter, we focus on 3D reconstruction of objects where the surface reflectance behavior might vary from diffuse reflectance to a reflectance behavior with both a diffuse and an additional specular component. For this scenario, we present a novel multi-camera, multi-projector super-resolution framework for structured light based 3D reconstruction which has successfully undergone peer review (see [WSRK11]). This system has been developed in the context of cultural heritage, where the focus is placed on completely automatic acquisition procedures and an as-accurate-as-possible, photo-realistic depiction of the object regarding both geometry and reflectance. This requires acquiring photos of the objects from a huge number of different viewpoints and under several different illumination conditions. For this reason, typical acquisition setups involve a rather expensive equipment. In our acquisition device, we use a multitude of cameras and projectors mounted on a hemispherical gantry above the object to be scanned. This approach allows the reconstruction of an object without moving either the object or the acquisition setup. Consequently, any registration of independent measurements is avoided. However, this kind of setup imposes severe restrictions on the type and placement of the projectors used to reconstruct the 3D geometry. Our contribution is dedicated to overcoming the limited resolution of the individual projectors which currently represent the limiting factor regarding the achievable resolution in multi-camera, multi-projector systems. The geometry acquisition pipeline described in the following sections has later been used as the prerequisite for accurate reflectance acquisition in [SWRK11].

After providing the motivation for our technique and discussing the contributions (see Section 3.1), we discuss the individual technical aspects required for an accurate reconstruction. These aspects include the used structured light encoding (see Section 3.2) as well as our approach to multi-projector super-resolution (see Section 3.3), our iterative robust bundle adjustment technique used for a selfcalibration of the acquisition device and the reconstruction of a point cloud (see Section 3.4), and the technique applied to reconstruct a closed surface mesh (see Section 3.5). This is followed by a description of the used acquisition device (see Section 3.6) and a discussion of the results achieved with our technique (see Section 3.7). Finally, Section 3.8 concludes this chapter with a short summary and a discussion of limitations to be considered by further research.

3.1 Motivation

As discussed in literature and reviewed in Chapter 2, structured light based techniques have been proven to be among the most accurate scanning techniques for diffuse surfaces and surfaces with both a diffuse and a specular reflectance component as long as the diffuse component is still sufficient to allow for a viewpoint-invariant decoding of the patterns. The underlying principle relies on the projection of illumination patterns with a unique decoding of the individual projector pixels onto the object. The resulting pattern on the object surface is observed by one or several cameras. Using such active illumination helps to establish active correspondences which can be used to triangulate the respective 3D surface point. Such actively generated correspondences allow a significantly more accurate reconstruction than correspondences obtained using passive acquisition techniques such as multi-view stereo techniques.

When using projector-camera systems for 3D reconstruction, one key observation can be identified in the fact that the resolution of currently available projectors is significantly lower than the one available in standard cameras. As a result, the footprint obtained by projecting a unique projector pixel extent onto the object surface usually covers several pixels in the camera which observes the scene. As active illumination relies on the unique encoding and decoding of the per-pixel projector illumination, this means that the resolution of the surface points obtained during the reconstruction process is limited by the projector resolution. In order to obtain a denser and more accurate reconstruction of the surface with its fine surface details, it is therefore mandatory to overcome this limitation.

To cope with the lower resolution of the projectors in comparison to the cameras, an additional mechanical system is used in [RSGS10] for lens-shifting, and highly accurate reconstructions of object surfaces have been reached. However, since we want to avoid the use of mechanical systems, we instead exploit the fact that we have multiple projectors within our multi-camera setup to increase the projector resolution. In particular, our approach to overcome the resolution limitations of the individual projectors is based on a novel super-resolution scheme introduced in this chapter. Sequentially illuminating the object using projectors at a sufficient number of different locations to cover the whole object surface with several structured light patterns allows to exploit the overlap of the patterns of projectors at different positions, which we observed for almost all of the surface points in our setup. We will demonstrate that, combined with an iterated bundle adjustment, these improvements increase the accuracy of the obtained point cloud significantly. In addition to this super-resolution technique for multi-camera, multi-projector systems, we also exploit using high dynamic range imaging to relax the assumptions regarding the reflectance behavior of the objects to be scanned, and even complicated objects exhibiting strong specularities can be scanned.

Employing structured light techniques within setups consisting of several projectors and cameras has been proposed in some previous methods. In recent years, a self-calibrating, multi-viewpoint approach to object acquisition, which is based on structured light and photometric methods, has been proposed in [AX08, AX10]. The 3D acquisition is performed by using a single projector placed at several positions or multiple projectors respectively as well as multiple cameras and by exploiting the fact that digital projectors can be simultaneously used as either active light sources or as virtual cameras. This is especially advantageous if only a few cameras are available. However, in contrast to our approach, Aliaga et al. [AX08, AX10] do not utilize the overlapping of different projected patterns on the object surface. For the purpose of refining both the projector pose and the 3D points, an iterative bundle-adjustment is applied in [AX08, AX10] in combination with an outlier rejection. Similarly, we also use an iterative bundle adjustment for the refinement of the calibration of our cameras but neglect the calibration of the projectors, since there is a sufficient number of cameras for establishing correspondences within the captured images.

In summary, the main contributions of this chapter are:

- a technique for the reconstruction of the full shape of an object in a single measurement via a multi-view, multi-projector structured light approach, eliminating the need for registering multiple measurements,
- a novel super-resolution scheme to overcome the limited projector resolution by combining overlapping structured light patterns,
- an optimized simultaneous automatic calibration of all cameras and the acquisition of the complete point cloud, using all available projector patterns, by a robust bundle adjustment on the whole dataset, and
- the use of high dynamic range imaging to increase the robustness of the reconstruction for objects with a more complex surface reflectance behavior.

3.2 Structured Light Encoding

Focusing on static scenes, we use a temporal encoding where multiple patterns are projected onto the object and the sequence of illumination values for a specific point across all of the patterns is used as the corresponding codeword. Furthermore, we only use binary patterns instead of gray-level or color patterns, which has the advantage of being more robust to noise. This is a crucial issue with respect to the acquisition of objects with complex reflectance behavior, and, as we are also interested in unique codewords, we use an encoding based on Gray code patterns [ISM84]. Gray code patterns enable a more robust decoding in comparison to simple binary code patterns. For adjacent codewords, Gray codes differ in only one bit. Since the decoding errors are most likely in the transition regions of the projected patterns, the fact that there is only one transition between adjacent codewords reduces the probability of errors. Furthermore, flips of these bits result in only small deviations as the decoded value differs only by one from the correct one. To further improve the robustness of Gray code decoding we employ an approach proposed in [Tro95], i.e. to use the completely white and the completely black pattern as well as the inverses of the previous patterns in order to avoid the determination of surface albedos. This technique was also used in [SS03] and [AX08]. Hence, the classification of whether a pixel is illuminated or not is carried out via comparing whether the pixel appears brighter when using the normal pattern or the inverse pattern.

3.3 Multi-Projector Super-Resolution

In order to increase the resolution of multi-camera, multi-projector structured light systems, we directly exploit the information contained in the overlaps of different pattern projections on the surface. Therefore, we have to consider a novel, modified encoding to generate super-resolution codewords (see Section 3.3.1) and also the respective decoding of these observed super-resolution codewords (see Section 3.3.2).

3.3.1 Super-Resolution Labeling

For the sets of projectors $\mathcal{P}$, cameras $\mathcal{C}$ and Gray code patterns $\mathcal{G}$, let $\mathcal{I} = \{I_{c,p,g} | (c, p, g) \in \mathcal{C} \times \mathcal{P} \times \mathcal{G}\}$ be the captured images. Thus, $I_{c,p,g} \in \mathcal{I}$ is an image (with domain $\Omega_c \subset \mathbb{R}^2$) captured by camera c, illuminated by projector p projecting pattern g.

By combining the $|\mathcal{G}|$ different pattern images for each pair of camera c and projector p, the Gray code can be decoded to labels $l \in [0, \dots, 2^{|\mathcal{G}|}]$. This defines a function $\chi_{c,p} : \Omega_c \to \mathcal{L}_{c,p}$, assigning a projector-dependent label $l \in \mathcal{L}_{c,p}$ to every image point $(x, y) \in \Omega_c$. Here, $\mathcal{L}_{c,p} \subset [0, \dots, 2^{|\mathcal{G}|}]$ is the set of labels that were actually decoded for camera c under illumination p. By inverting the labeling function $\chi_{c,p}$, we can identify the image regions $\mathcal{R}_{c,p,l} \subset \Omega_c$, which have been labeled with l.

The super-resolution labeling function is defined as

$$\begin{aligned} \xi_c : \qquad\qquad \Omega_c &\to \mathfrak{L}_c \\ (x, y) &\mapsto (\chi_{c,p_1}(x, y), \dots, \chi_{c,p_{|\mathcal{P}|}}(x, y)) \end{aligned} \tag{3.1}$$

assigning to every image point (x, y) a projector-independent label

$$L = (L_1, \dots, L_{|\mathcal{P}|}) \in \mathfrak{L}_c \subset \mathcal{L}_{c,1} \times \mathcal{L}_{c,2} \times \dots \times \mathcal{L}_{c,|\mathcal{P}|} \tag{3.2}$$

with $L_i = \chi_{c,p_i}(x, y)$. L can be understood as the tuple concatenating the labels of the individual projectors. Analogical to the projector-dependent label, we can invert ξ and get the image regions $\mathfrak{R}_{c,L} \subset \Omega_c$ which have been labeled with L.

As shown in Figure 3.1, the region $\mathfrak{R}_{c,L}$ is the intersection $\bigcap_{i=1}^{|\mathcal{P}|} \mathcal{R}_{c,p_i,L_i}$, resulting in a usually more precise localization in the image for the super-resolution labels when compared to the original labels.

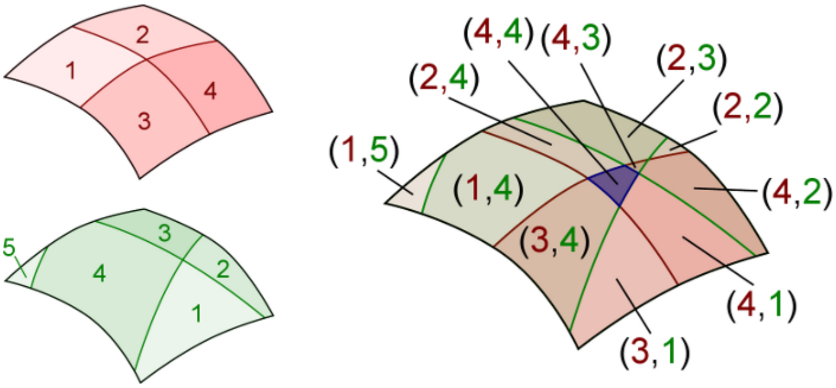


Figure 3.1: *Super-resolution scheme for structured light: By combining the two sets $\mathcal{L}_1$ (red) and $\mathcal{L}_2$ (green) of labels obtained from two independent projectors into super-resolution labels $\mathfrak{L}$, much smaller regions on the surface can be identified uniquely. For example, the highlighted blue region $\mathfrak{R}_{(4,4)} = \mathcal{R}_{1,4} \cap \mathcal{R}_{2,4}$ is much smaller than both of the intersected regions.*

To increase clarity and shorten the notation, we will dismiss the camera and projector indices in the remainder of this chapter, unless they are needed. Please keep in mind that the super-resolution labels are always defined per camera and the projector dependent labels per pair of camera and projector.

3.3.2 Robust Matching of Codewords

For a robust matching, we extend the codomain of the projector-dependent labeling function χ to $\mathcal{L}' = \mathcal{L} \cup \{\perp\}$, including $\perp$ to express *undefined*. The symbol $\perp$ is used for all image points (x, y) for which the label could not be decoded, either because they were not lit by the projector due to occlusions by the geometry or because they could not be classified reliably as shown in Figure 3.2.

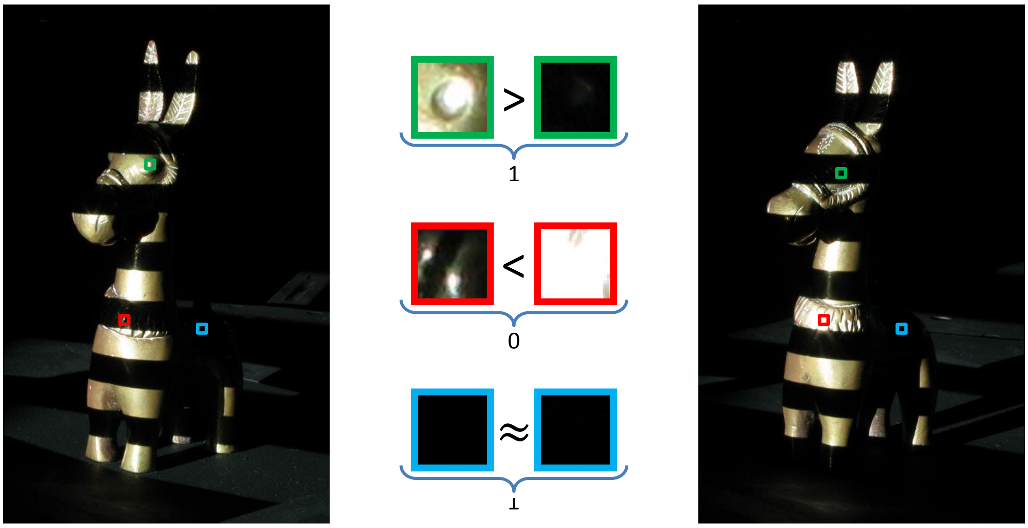


Figure 3.2: *Illustration of the derivation of structured light encodings: If a pixel value is significantly brighter under illumination with a certain stripe pattern than the respective value under illumination by the corresponding inverse structured light pattern (marked in green), the respective bit in the encoding is set to* 1*. In contrast, when the pixel value is darker than its corresponding value observed under illumination by the inverse pattern (marked in red), the bit in the encoding is set to* 0*. If the values under the structured light pattern and its inverse are close to each other, no reliable decision can be made for the respective bit in the decoding (marked in blue) which we mark by using* $\perp$ *in the encoding.*

As a consequence, the super-resolution labels are extended as well to $\mathfrak{L}'$, with the tuple $L \in \mathfrak{L}'$ containing $\perp$ at entries where the label for the corresponding projector could not be decoded. Please note that this will introduce a new inverse labeling function

$$\begin{aligned} \xi'^{-1}: \quad & \mathfrak{L}' \to \Omega \\ & L \mapsto \{(x,y) \in \Omega | \forall i \in [1, \ldots, |\mathcal{P}|]: \; L_i = \perp \vee L_i = \chi_{p_i}(x,y)\} \end{aligned} \tag{3.3}$$

for a super-resolution label $L = (L_1, L_2, \ldots, L_{|\mathcal{P}|})$. This means that $\mathfrak{R}'_L$ is in fact the intersection of the regions of all projections i for which a sub-label L_i could reliably be identified, ignoring the indeterminable region of $\perp$ in projections with undefined entries.

It might occur that a normal label l projected by projector p_k, which was correctly decoded in camera c_i, was not reliably classified in camera c_j. Therefore, all super-resolution labels $\{(L_1, L_2, \ldots, L_{|\mathcal{P}|}) | L_k = l\} \subset \mathfrak{L}'_{c_i}$ of camera c_i that contain l at the k^{th} index cannot be found in $\mathfrak{L}'_{c_j}$. However, if the labels in all other projections were classified correctly in both cameras, we would still like to use this partial match, providing us with a reasonably precise region $\bigcap_{i \neq k} \mathcal{R}_{p_i}$ (see Figure 3.3 for an example). Thus, we introduce a partial matching function P_c^M, that will robustly match any super-resolution label in any camera with all super-resolution labels detected in camera c.

$$\begin{aligned} P_c^M: \quad & \bigcup_{c' \in \mathcal{C}} \mathfrak{L}'_{c'} \to 2^{\mathfrak{L}'_c} \\ & (L_1, \ldots, L_{|\mathcal{P}|}) \mapsto \{(M_1, \ldots, M_{|\mathcal{P}|}) \in \mathfrak{L}'_c | \; \forall i: L_i = M_i \vee L_i = \perp\} \end{aligned} \tag{3.4}$$

Consider a super-resolution label $L \in \mathfrak{L}'$ that was incompletely identified in camera c_j and the matched set of super-resolution labels $P_{c_i}^M(L)$ in camera c_i. The region in c_i corresponding to the region of L in c_j is then determined by $\bigcup_{M \in P_{c_i}^M(L)} \xi'^{-1}(M)$. For a sketch we refer to Figure 3.3. Please note that even though the position of a region belonging to a label is the actual information needed in the multi-view-triangulation, its size also yields important information about the accuracy, which will be used in the bundle adjustment (see Section 3.4) as well as the surface reconstruction (Section 3.5). This fact is especially important in our case, since we perform a triangulation with super-resolution labels of very different sizes and therefore different accuracies.

3.4 Iterated Robust Bundle Adjustment

For the reconstruction of an accurate point cloud, the correct calibration of the cameras is of utmost importance. Furthermore, in our setting this calibration

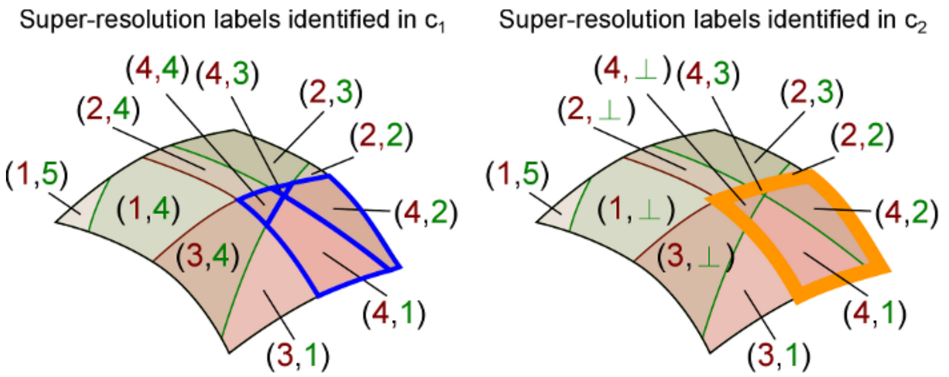


Figure 3.3: *Toy example for robust matching: The label* 4 *for projector* 2 *(green) could not be reliably classified in camera* c_2 *(right) and was substituted by* $\perp$*. Yet, a partial matching for the query-super-resolution-label* $(4, \perp)$ *(orange) in camera* c_1 *will find* $P^M_{c_1}((4, \perp)) = \{(4,1), (4,2), (4,3), (4,4)\}$ *(blue) and a correspondence between* c_1 *and* c_2 *on a coarser super-resolution level can still be established. Please note that, even though only a small patch in the right image is labeled as* $(4, \perp)$*, the inverse of this label with uncertainty* $\xi^{-1}((4, \perp))$ *is in fact the complete orange region.*

can afterwards also be used for the reflectance acquisition, as a good camera calibration is also necessary there. Determining the calibration of a camera c, i.e. the projection Π_c which maps a point onto its image in the image plane Ω_c, from correspondences between 3D points $X \in \mathbb{R}^3$ and their known projected positions $x_c \in \Omega_c$ is a well-studied problem. For the case of multiple cameras, the method of *Bundle Adjustment* (BA) [TMHF00] is a widely established solution based on statistical optimization. The key idea is that the reprojection error $E = \sum_{\hat{X} \in \mathcal{X}} \sum_{c \in \mathcal{C}_{\hat{X}}} d(\hat{\Pi}_c(\hat{X}), x_c)$ of the 3D points in all cameras $c \in \mathcal{C}$ is minimized with regard to a metric d by estimating an optimal projection (i.e. camera calibration) $\hat{\Pi}_c$ for all cameras $c \in \mathcal{C}$ and a point cloud $\mathcal{X} = \{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_n\}$. Here, $\mathcal{C}_X$ denotes the set of cameras in which X was visible, i.e. the super-resolution label $L^X \in \mathfrak{L}'$ corresponding to this point on the object's surface was identified. This means that the projection of X has to lie within the region $\mathfrak{R}'_{c,L^X} = \xi'^{-1}_c(L^X)$ for each camera $c \in \mathcal{C}_X$. To quantify this knowledge, we define the center of mass $\bar{x}$ of $\mathfrak{R}'_{c,L^X}$ as x_c and utilize the Mahalanobis distance between x_c and the projected point $\hat{\Pi}_c(\hat{X})$ as the reprojection error metric d. We weight the Mahalanobis distance with the covariance matrix

$$\sigma_{ij} = \int_{x \in \mathfrak{R}'_{c,L^X}} (x_i - \bar{x}_i)(x_j - \bar{x}_j)\, dx \tag{3.5}$$

of the region.

The original approach is only feasible for a relatively small set of very confident correspondences, though. With *Sparse Bundle Adjustment* (SBA) [HZ04, LA09] a faster solution is available, making the processing of larger sets of correspondences possible.

However, the problem remains that (S)BA is not very tolerant with respect to outliers, since existing algorithmic implementations are at risk of finding only local minima. Similar to [AX08, AX10, SSS08], we employ a RANSAC strategy [FB81] to sort out outliers beforehand and provide a good initial estimation of 3D points and camera calibration matrices. For this, we use two separate RANSAC passes, one to determine the initial 3D point cloud and one to provide an initialization for the camera calibrations. First, each point X is computed by repeatedly drawing a small subset $C \subseteq \mathcal{C}_X$ of cameras randomly and checking its reprojection error E_1 in all cameras $\mathcal{C}_X$. Cameras with errors below a certain threshold are considered to be inliers. In a second pass, a robust initialization for the camera projections Π_c for all cameras $c \in \mathcal{C}$ is computed the same way by using a small number of reconstructed points $X' \subseteq \mathcal{X}_c = \{X | c \in \mathcal{C}_X\}$ that were seen by camera c. Once again, the inliers are determined by thresholding a reprojection error E_2, but this time for all the points in $\mathcal{X}_c$. In both cases, the solution with most inliers is regarded as a good initialization for SBA and recomputed using all inliers in a final refinement step. Furthermore, for a robust computation, it is important to remove outliers also during the computation of the SBA. We thus discard bad correspondences with regard to E_2.

After the SBA, improved camera calibration matrices are available. Thus, by repeating the RANSAC process, a better outlier removal and with it a better initialization for the SBA is possible. This will in turn result in a more precise calibration. Thus, like in [AX08, AX10], we iterate this procedure until the reprojection errors E converge.

Please note that our iterative approach requires a first (coarse) calibration of the cameras. We will address this issue in Section 3.6.

3.5 Surface Reconstruction

Since for many applications a polygon mesh is a better suited representation than a point cloud, we apply the *Poisson Reconstruction* introduced in [KBH06], which generates closed surfaces. Furthermore, this method is able to robustly handle remaining outliers, noise and holes which occur in regions not lit by any projector. For the surface generation process, the Poisson Reconstruction requires oriented point clouds, i.e. knowledge about the normals for every point. Thus, an approximation of these normals is typically computed, where the k-nearest neighbors in space are considered and a quadric or plane surface is fitted. When using this strategy in the presence of thin parts of the object, points from both sides of the surface will be considered as neighbors, which results in an erroneous estimation of the normals. Instead, we exploit the fact that points lit by the same projector are usually not on different sides of thin object parts except for very degenerated theoretical configurations, which allows for a robust estimation of the normals.

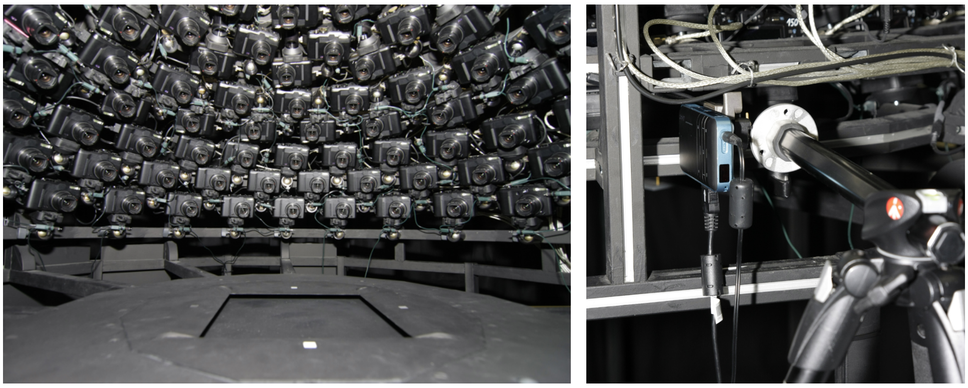
Since our proposed super-resolution labeling approach leads to reconstructed points with a considerable difference in accuracy, this should be taken into account during the surface reconstruction. A confidence ellipsoid in $\mathbb{R}^3$ for every reconstructed point can be computed by back-projecting the known covariances σ_{ij} (see Equation (3.5)) in the image domain for every region that was involved in the triangulation [HZ04]. As the implementation proposed in [KBH06] only supports a single confidence value per point, we suggest to project the ellipsoid onto the normal direction to get a good estimate of the confidence of the points, although we did not investigate this option in the scope of this work.

Alternatively, the more recent *Floating Scale Surface Reconstruction* technique published in [FG14] might be applied instead of the Poisson Reconstruction. In this method, the scale is considered for each 3D point that has been reconstructed via triangulation. It might be possible to derive a scale parameter from the aforementioned confidence parameter, which might be followed in the scope of further investigations.

While we expect the reconstruction results to become more accurate based on such modifications, the high-frequency noise inherited in triangulation-based methods is likely to remain and, hence, fine surface details might still not be accurately reconstructed but rather smoothed by the surface reconstruction technique. Instead, we will focus on complementing the triangulation-based information obtained by using our structured light system with information about the locally measured surface normals in the following chapter. Such normal-based geometry acquisition techniques typically succeed in accurately reproducing fine surface details but suffer from a low-frequency drift which can be compensated by a combination of a normal-based technique and a triangulation-based technique.

3.6 Acquisition Setup

For our experiments, we extended the multi-view setup described in [MBK05] by placing projectors at several positions and using cameras with a higher resolution.



(a) Camera hemisphere and sampleholder.

(b) One projector on a tripod.

Figure 3.4: *Our experimental setup.*

This setup has originally been designed for the purpose of reflectance acquisition of flat material samples and extended for the acquisition of photo-realistic 3D models of objects by accurately capturing surface geometry in addition to reflectance behavior. Within the proposed setup, 151 Canon PowerShot G9 cameras with a resolution of 12 Megapixels are mounted uniformly distributed on a hemispherical gantry with a diameter of approximately 1.3m (see Figure 3.4a). We simulated the use of multiple projectors by placing one projector, which was mounted on a tripod, at different positions (see Figure 3.4b) in the scope of the publication of this approach in [WSRK11].

The number of projector placements $|\mathcal{P}|$ and used cameras $|\mathcal{C}|$ varies for the different results presented. Please see Table 3.1. To acquire the objects from all sides at once, the projector placement has to be chosen in such a way that every point on the surface is lit at least once. A good configuration is to place the projector at several azimuthal steps with a declination of $15°$, and use two or three additional positions from about $70°$ declination if necessary. There are certain constraints for placing the projectors. On the one hand, every point on the surface should be illuminated by at least one projection (for the utilization of the super-resolution technique at least two projections) and the patterns should be as fine as possible. On the other hand, the object should not be occluded by the projector in any camera. We therefore choose positions between the cameras from where the object is illuminated. This requires a rather small projector to fit into the spaces and a minimum distance of about 60cm. We employed an Acer C20 Pico LED-Projector with a resolution of 848×480 pixels which results in $|\mathcal{G}| = 19$ patterns, i.e. $2 \cdot |\mathcal{G}| = 38$ patterns including the inverse patterns, being sufficient to utilize the projector resolution completely. Meanwhile, the setup is equipped with nine LG HS200G projectors (800×600 pixels, LED-DLP, 200 lm) to avoid any manual effort during the acquisition. In more detail, six of these projectors are placed at an inclination angle of $\theta \approx 82.5°$ with an even spacing of $\Delta\varphi = 60°$ and three projectors are located at $\theta \approx 17°$ with $\Delta\varphi = 120°$.

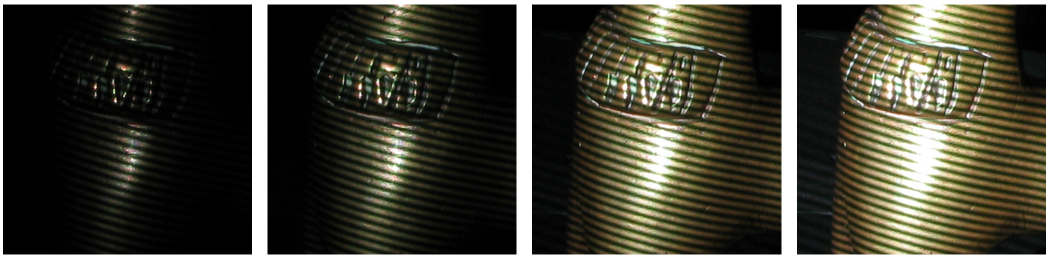
As we are also interested in acquiring the geometry of objects with more complex reflectance properties, we take HDR sequences and, hence, have to acquire several exposure steps for every captured HDR image. The number of exposure steps N is chosen according to the dynamic range of the material of the object surface which is typically based on the judgement of the acquisition expert. However, this might also be performed automatically as discussed in Chapter 9.

Altogether, we acquire a total number of $|\mathcal{C}| \times |\mathcal{P}| \times 2 \cdot |\mathcal{G}| \times N$ images. Due to the massive parallelization, the system is able to capture 151 images simultaneously. The complete acquisition time for the Cueball dataset (see Table 3.1) consisting of $151 \times 10 \times 38 \times 3 = 172{,}140$ images was approximately 2 hours, including the time for repositioning the projector.

During the reconstruction of the object, we find a precise camera calibration using SBA (see Section 3.4). However, we do need a decent initial camera calibration. For this, we use a coarse camera calibration computed for the multi-camera setup, that is performed beforehand using a calibration target. Note that this step does not need to be carried out for every measurement, but only when the setup is changed.

3.7 Experimental Results

In order to evaluate our approach, we acquired scans of two objects. We demonstrate the feasibility of our method with a challenging glossy metallic donkey statue. To handle the specular highlights, several differently exposed images had to be taken and combined to an HDR image. As illustrated in Figure 3.5, the structured light pattern can be correctly observed in highlight regions for a rather short exposure time, whereas the pattern cannot be reliably observed on the remaining surface parts due to underexposure. In contrast, for longer exposure times the stripe pattern can be clearly decoded in the previously underexposed regions, but not reliably be observed in the highlight regions anymore due to overexposure. Please see Figure 3.6 for the obtained reconstruction of the complete model. We also acquired the donkey statue with a NextEngine laser scanner, which is a commercial low cost acquisition system and widely used for acquiring 3D objects, particularly in the field of cultural heritage. A comparison with the reconstruction obtained with our technique is shown in Figure 3.7. Clearly, our method is also capable of reconstructing the fine engravings on the forehead of the donkey which are smaller than the finest projected patterns, where the laser scanner has insufficient resolution. Since no sufficiently precise ground truth dataset is available for the donkey figurine, it is not possible to determine the accuracy of the resulting reconstruction by measuring the distance to the ground truth. Nonetheless, to get a rough estimate of the noise level, we compute the Hausdorff distance to the result obtained after applying Poisson Reconstruction, which is a smooth approximation of the real surface and obtain an accuracy of 144μm. More details are given in Table 3.1.



(a) 125ms **(b)** 500ms **(c)** 2000ms **(d)** 4000ms

Figure 3.5: *An HDR sequence of the same structured light pattern on the donkey example. Obviously, there is a need to consider several exposure times: The lines through the highlight, for example, are clearly visible in image (a). However, they cannot be distinguished in image (d) due to overexposure. The same holds true for underexposed parts. Note that there is no single exposure step that shows all patterns on all parts of the surface.*

Furthermore, we determine the precision of the reconstruction using a white billiard cueball, which we consider to be a good reference object, almost resembling a perfect sphere. For this purpose, we performed a least squares fit of a sphere to all points that we manually identified as part of the cueball (in contrast to the ground plane, on which the object was placed) and measured the distance of every point in the reconstructed point cloud to the surface of this sphere. The accuracy of the reconstructed point cloud consisting of 4,753,069 3D points can be seen in a RMS of 23.3μm as mentioned in Table 3.1. In Table 3.2, we evaluate the improvement over the iterations of the robust iterated SBA approach and compare the achievable precision with super-resolution labels against projector-dependent labels.

To get an impression of the underlying principle, we refer to Figure 3.8. Here, two sets of projector-dependent labels and the resulting set of super-resolution labels obtained by combining them are shown. Obviously, the super-resolution labels are much smaller and, thus, provide a better localization. To quantify this, in Figure 3.9 the histogram of the label sizes is depicted. As expected, the number of super-resolution labels is much higher and they are distributed in smaller regions.

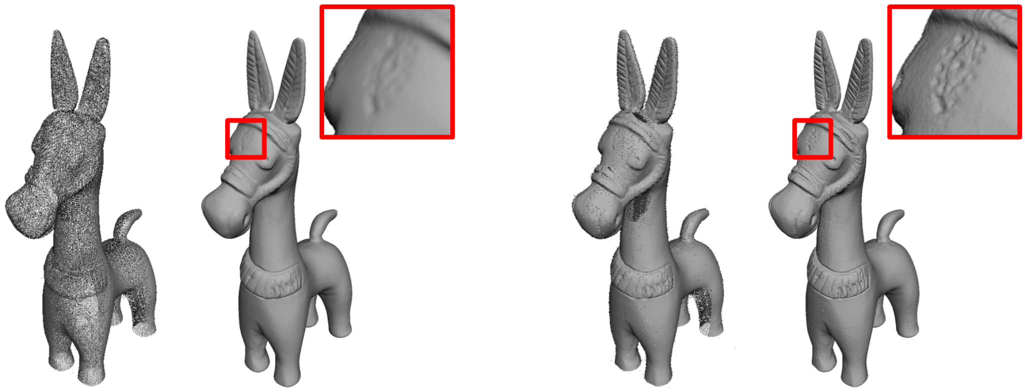


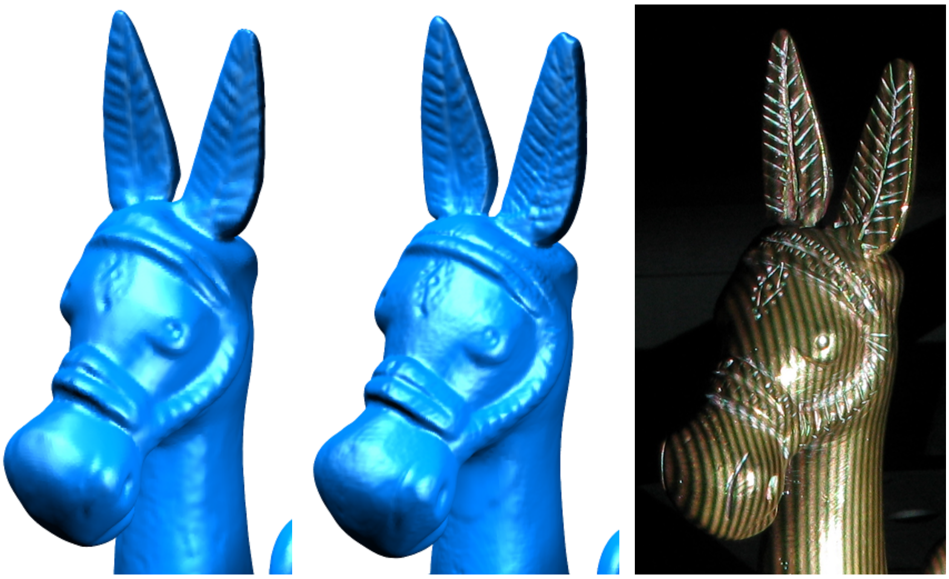
Figure 3.6: *Comparison of the reconstruction results obtained with the conventional Gray-code based structured light approach and the proposed super-resolution approach: From left to right, there are visualizations of the point cloud obtained using the conventional Gray-code based structured light approach, the result after applying Poisson Reconstruction [KBH06] to this point cloud as well as the point cloud created with the proposed super-resolution approach and the respective result after applying Poisson Reconstruction [KBH06]. The point cloud obtained with the proposed super-resolution approach is significantly denser and allows a more accurate reconstruction of fine surface details. In the region of the engraving on the forehead of the figurine, the resolution achieved with the conventional structured light approach is about* $204\mu m$*, while a resolution of about* $60\mu m$ *is achieved with the proposed super-resolution approach.*

The minor difference in the distribution of the projector-dependent labels is due to the placement of the projectors.

Furthermore, Figure 3.10 shows a direct comparison of a photo taken from a frontal view with respect to the respective object and the corresponding reconstruction for the donkey figurine and two further figurines.

3.8 Conclusions

In this chapter, we have presented a super-resolution approach based on structured light to significantly improve the resolution of multi-projector structured light systems. By exploiting the placement of several projectors within the setup in order to cover the whole object surface with light patterns, we utilized overlapping pattern projections for uniquely encoding much smaller regions on the surface. Within our experiments, we tested the performance of the proposed technique using several objects with challenging reflectance behavior and demonstrated that the proposed



(a) Laser scan. **(b)** Super-resolution. **(c)** Light pattern.

Figure 3.7: *Comparison of the results obtained when applying the Poisson Reconstruction [KBH06] to two point clouds: (a) A laser scan performed with the commercially available NextEngine Scanner and (b) a point cloud obtained by our proposed super-resolution approach. Using five different projector positions, we are able to reconstruct even fine surface details, which are smaller than the finest projected patterns. For a comparison, see (c), which shows the finest vertical pattern of the second projector.*

technique is capable of a dense and accurate reconstruction. Our technique has become one of the indispensable components in the subsequently published reflectance acquisition devices presented in [SWRK11, SSWK13, SSW^{+}14] which certainly belong to the most accurate devices that are currently available.

So far, we used a simple threshold for classifying the pixels into “illuminated” and “not illuminated”. In future work, our structured light decoding might be extended towards an even more robust classification for reflections such as the one proposed in [XA09], which is based on separating the direct and global components as proposed in [NKGR06]. Furthermore, our approach is still limited to the resolution of the employed cameras as we currently only detect the stripe transitions on a per-pixel level within the image. Hence, the overall quality of the reconstruction is expected to be improved by also considering a subpixel accurate localization of the transitions between neighboring stripe projections on the object surface, e.g. as it has been applied in [Tro95]. As a consequence, the combination of both a subpixel accurate detection of the stripe transitions and our proposed multi-projector super-resolution could be a promising avenue of future research.

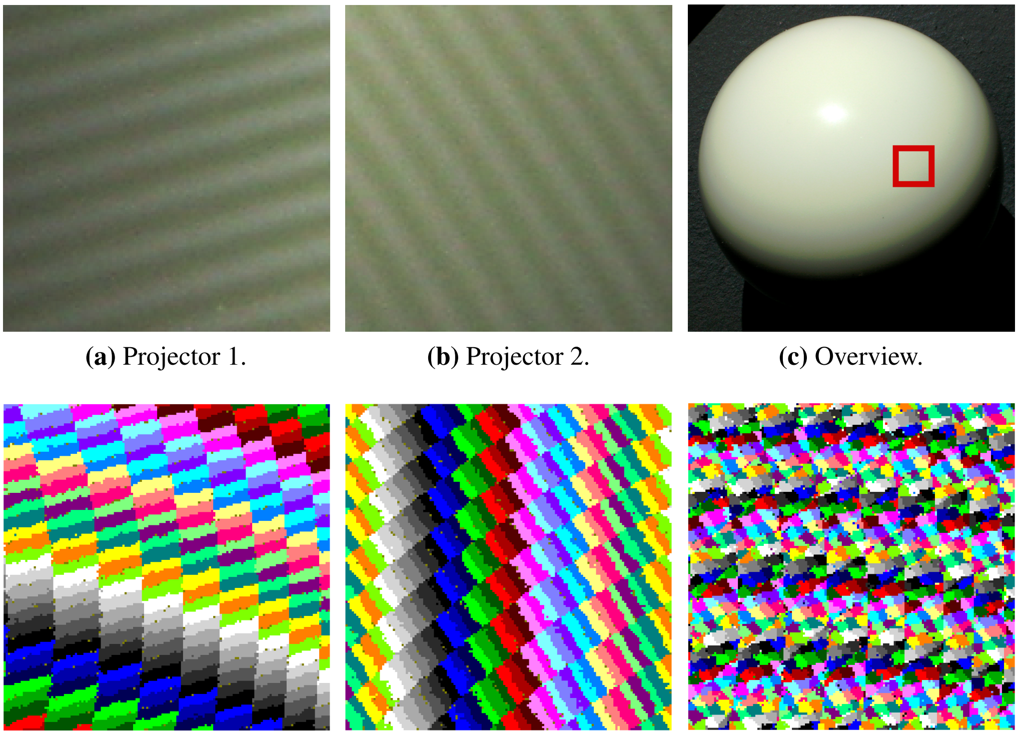
However, one of the key observations obtained during the development of our technique was the insight that triangulation-based reconstruction methods strongly rely on accurately localized correspondences and an exact calibration of the setup and suffer from noise if there are inaccuracies regarding these aspects. Therefore, we instead focused on complementing the 3D acquisition technique presented in the scope of this chapter with a reconstruction technique based on normal information to further improve the reconstruction quality as described in Chapter 4.

Table 3.1: *Details for our test cases. The size of the donkey was measured in the point cloud. For the size of the cueball, we used the manufacturer description. The projector resolution was determined in images which were taken close to the projector on a surface facing the camera. The image resolution was determined in the topmost camera.*

	Donkey	**Cueball**
$\lvert\mathcal{C}\rvert$	150	51
$\lvert\mathcal{P}\rvert$	5	10
$\lvert\mathfrak{L}'\rvert$	$5{,}039{,}375$	$37{,}800{,}626$
size	$10 \times 4 \times 17\text{cm}^3$	$\varnothing 6.02\text{cm}$
projector resolution	$\approx 500\mu\text{m}$	$\approx 500\mu\text{m}$
camera resolution	$\approx 100\mu\text{m}$	$\approx 47\mu\text{m}$
accuracy (RMS)	$\approx 144\mu\text{m}$	$23.3\mu\text{m}$
points	$562{,}798$	$4{,}753{,}069$

Table 3.2: *Accuracy evaluated on the cueball dataset: Super-resolution labels from* 10 *projectors and* 51 *cameras (RMS 51,...),* 10 *projectors and* 11 *cameras (RMS 11,...) and the concatenation of individual point clouds of projector-dependent labels with* 51 *cameras (RMS PDL,...).* 3 *intermediate iterations of the robust bundle adjustment have been performed in all cases. All numbers are specified in µm and normalized to the manufacturer specification of the size of the cueball (*$\varnothing 6.02$*cm).* RMS *is the root-mean-square error measure,* mean *is the average error and* max *denotes the maximum error. We need to stress that we did not manually remove any outliers but only the ground plane. Please note that for the iterations only a subset of points was used in the SBA.*

Iteration	RMS 51	mean 51	max 51	RMS 11	mean 11	max 11	RMS PDL	mean PDL	max PDL
init.:	127.7	88.1	11,573.1	167.2	117.7	11,880.6	119.7	55.4	12,135.4
iter. 1:	45.7	29.7	968.4	121.4	85.7	3,317.2	6,743.0	5,315.1	26,683.8
iter. 2:	25.6	18.5	778.0	101.7	77.9	472.4	5,724.2	4,472.5	23,337.8
iter. 3:	24.2	17.9	449.7	99.3	75.4	688.9	5,203.0	3,975.0	25,334.9
final:	23.3	17.5	2,856.6	96.7	74.2	463.1	126.0	92.9	5,240.6



(a) Projector 1. **(b)** Projector 2. **(c)** Overview.

(d) Labels 1. **(e)** Labels 2. **(f)** Super-resolution labels.

Figure 3.8: *A super-resolution label example using two projectors on the topmost camera of the cueball dataset. In the images (a) and (b), a* 150×150 *pixels cutout of the input configuration is shown which is also depicted in image (c) with the finest horizontal pattern (one stripe is approximately* $500\mu m$ *wide) projected by projector* 1 *and* 2 *respectively. The area of the cutout corresponds to a square of about* $7mm \times 7mm$*. In the images (d) and (e), the decoded patterns for the single projectors are shown. To facilitate the distinguishing of the labels, the color-palette was limited to* 34 *colors and is repeated. The visualized labels are unique and do not repeat themselves. In image (f), the super-resolution labels are shown that are obtained by combining the two projectors.*

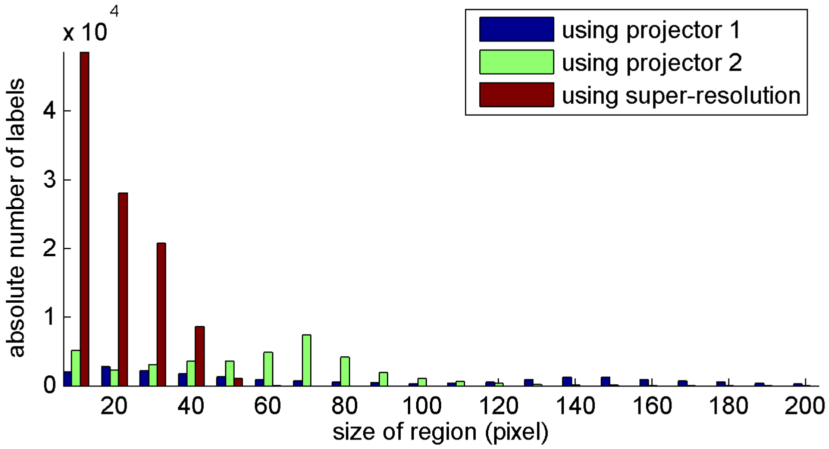


Figure 3.9: *Example distribution of label sizes obtained on the full cueball images from the example in Figure 3.8. The X-axis denotes the size of a region in pixels (10-pixel bins) and the Y-axis denotes the number of labels whose regions lie within that size. Our proposed super-resolution technique (red bars) creates a considerably larger number of much smaller labels. In this case, the super-resolution labels were constructed using only the two shown projectors. However, using additional projectors will further decrease the size of the regions.*

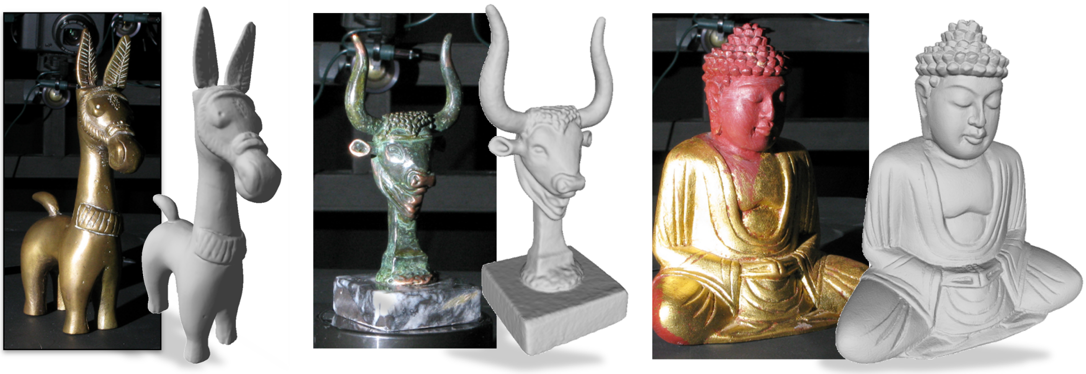


Figure 3.10: *Reconstruction results for several objects.*

CHAPTER 4

FUSING STRUCTURED LIGHT CONSISTENCY AND HELMHOLTZ NORMALS

Our acquisition technique based on a super-resolution scheme for multi-camera, multi-projector structured light systems as introduced in the previous chapter is capable of densely reconstructing the full 3D shape of objects in the range from diffuse surface reflectance to specular surfaces with a sufficient diffuse component. Similar to other techniques that are also based on the triangulation of point correspondences, there might still be some visible noise in the reconstructed model as a result of inaccurately localized correspondences and a non-ideal calibration. Accordingly, there might still be a lack of accuracy in the reconstruction of even finest surface details such as fine scratches or engravings which represents the major limitation of the structured light based approaches.

In the scope of this chapter, we present a novel 3D reconstruction approach which combines a structured light based consistency measure with dense normal information obtained by exploiting the Helmholtz reciprocity principle and has successfully undergone peer review (see [WRO+12]). This combination compensates for the individual limitations of techniques providing normal information, which are mainly affected by low-frequency drift, and of those techniques providing positional information, which are often not well-suited for recovering fine details. In order to obtain Helmholtz reciprocal samples, we employ a turntable-based setup. Due to the reciprocity, the structured light directly provides the occlusion information needed during the normal estimation for both the cameras and light sources. This allows performing the reconstruction by solving one global variational problem which integrates all available measurements simultaneously, over all cameras, light source positions and turntable rotations. For this purpose, we employ an octree-based continuous min-cut framework in order to alleviate metrification errors while maintaining memory efficiency. The performance of our approach is evaluated both on synthetic and real-world data where highly accurate reconstructions have been achieved.

After discussing the main contributions of our technique presented in the scope of this chapter (see Section 4.1), we provide an overview of different approaches to combine different cues for 3D reconstruction that have been presented in literature (see Section 4.2). This is followed by a description of our variational framework for combining a novel structured light based consistency measure with estimated Helmholtz normals (see Section 4.4). After providing implementation details (see Section 4.5) and details about the used acquisition setup (see Section 4.3) we show results achieved with our proposed technique on different datasets (see Section 4.6) and provide concluding comments (see Section 4.7).

4.1 Motivation

Due to their impressive reconstruction accuracies, laser scanners or structured light systems still represent widespread standard solutions to acquire the 3D shape of objects. As discussed in the previous chapter, the accuracy of structured light systems is mainly limited by the resolution of the projectors. Investigations that focus on overcoming this limitation include the techniques presented in [RSGS10] and the previous chapter. While using lens-shifted structured light has been investigated in [RSGS10], our technique described in the previous chapter exploits combining codeword information within a multi-camera and multi-projector setting to overcome these limitations. However, a still remaining major drawback of all methods relying on the triangulation of feature correspondences is that this triangulation entails high-frequency noise in the final reconstruction. This noise is mainly due to inaccurate feature localization or due to calibration inaccuracies. Therefore, such 3D reconstructions based on triangulation either suffer from noise or over-smoothing. As a result, fine details might not be captured. In contrast, a reconstruction based on normal integration would be capable of preserving high-frequency surface details but is prone to low-frequency drift due to the accumulation of errors. Combinations of these two reconstruction principles have been proposed to overcome the respective problems. In the following sections, we will investigate the combination of information derived using a structured light system with normal information estimated by utilizing the Helmholtz reciprocity principle. Both techniques make only modest assumptions about surface reflectance. As a result, they can be used to capture nearly all opaque objects; the only notable exception being highly specular objects, such as mirrors. Therefore, our geometry acquisition method described in the scope of this chapter competes with the methods designed for rough surfaces with diffuse or near diffuse surface reflectance and glossy surfaces with both a diffuse component and a specular component. Additionally, the combination of structured light information with Helmholtz normal information is especially useful in the setting of photo-realistic reproduction, since the images needed for reflectance reconstruction can also be used for the Helmholtz normal estimation. The use of structured light directly provides information about occlusion and shadowing that can be utilized in the Helmholtz stereopsis.

Most existing reciprocal setups, i.e. devices for which the view- and light-direction can exactly be exchanged, rely on moveable cameras and light sources. Instead, we employ a device with fixed light sources, cameras and projectors, using a turntable as the only moving part. This simplifies calibration considerably and reduces hardware complexity. By using a symmetric layout of the light sources, such a setup can be utilized to obtain pairs of Helmholtz reciprocal samples (see Figure 4.1).

Since the turntable rotates during the acquisition, the projected patterns move over the surface of the object. Therefore, a direct triangulation of codewords can only be performed per rotation. Instead of using all cameras that have seen one location on the surface to compute a single consistent point, an independent triangulation is performed for each rotation, resulting in multiple potentially contradicting solutions. Regaining the fine surface details from the resulting noisy point cloud, however, is a non-trivial task.

To overcome this limitation and obtain one globally consistent reconstruction integrating the information over all rotations, we use a variational approach which combines a consistency term derived from our structured light technique with Helmholtz normals. We solve the respective optimization problem via an octree-based continuous min-cut framework, which is memory-efficient and alleviates metrification errors. To compensate for the discretization artifacts from the min-cut, a smooth signed distance function is then computed from the resulting binary labeling, again taking the reconstructed normals into account. Finally, the reconstruction result is derived from this smooth signed distance.

In summary, the key contributions presented in this chapter are:

- a turntable-based symmetrical setup that allows to acquire reciprocal image pairs.
- a structured light based consistency measure which allows to combine several structured light measurements although the object was moved with respect to the projector,
- a variational reconstruction technique combining positional information obtained from a structured light technique and normals obtained from Helmholtz reciprocity,
- a memory-efficient octree-based continuous min-cut solver, and

- a final refinement step based on a smoothed signed distance function derived from the min-cut and the Helmholtz normals.

4.2 Combinations of Complementary Cues for 3D Reconstruction

In order to overcome the limitations of triangulation-based techniques, several investigations focus on fusing consistency or triangulated positional information with complementary visual cues such as normal orientation. Many approaches employ deformable models to guide an iterative surface evolution by optimizing a cost function to enforce consistency. For this purpose, the surface is often represented in a volumetric way [ES04, JCYS04] or by polygonal meshes [NRDR05, BCJS06, EVC08, LTBEB10, AX10, DPB10, DP11, WLDW11]. These methods rely on a good initialization and evolve the surface via a variational problem using gradient flows. However, such surface evolution techniques can get stuck in local minima. In addition, mesh-based techniques are not capable of handling topological changes of the surface geometry such as self-intersecting regions in the mesh. This can be overcome by hybrid approaches [YY11].

In contrast, global optimization strategies such as graph cuts [SP05, HK06, YAC06, VHTC07, LBN08, HMJI09] or convex optimization [KPC10] overcome the local minima problem by considering a cost function defined on the whole volume for surface extraction. Due to their discrete nature, graph cut approaches necessitate the use of high-resolution volumetric grids to obtain high-quality reconstructions. In addition, a high grid-connectivity is desirable to reduce metrification errors. Incorporating these aspects comes at the expense of a drastically increasing memory consumption. This problem can be alleviated to some extent by employing adaptive grid structures such as octrees. Convex optimization, in contrast, does not suffer from metrification errors. Thus, shifting variational problems from the discrete to the continuous domain has only recently gained attention in literature [CK11, KKH$^+$11, YBT10] although e.g. the continuous min-cut and max-flow formulation has already been introduced in the 1980s [Str83, Str10]. Although these techniques are more memory efficient, storing the data on a regular grid still imposes severe resolution restrictions. In this chapter, we adapt the continuous min-cut technique proposed in [YBT10] to an octree structure.

For a reliable reconstruction of fine surface details, the accuracy of the normals is a crucial factor. However, obtaining accurate normals for the wider range of non-Lambertian opaque objects is a challenging task. Therefore, several techniques rely on simplifying assumptions such as either Lambertian or purely specular surface reflectance, which severely limits the range of objects that can be acquired. In contrast, obtaining accurate normals for opaque objects has been investigated by exploiting the duality between cameras and light sources via the Helmholtz reciprocity [ZBK02], which requires an accurate co-location of the cameras and the light sources. The only assumption is that the surface materials in the scene have a reciprocal bi-directional reflectance distribution function (BRDF). Recently, this principle has also been used in a multi-view setting in [DPB10].

In our technique, we use the combination of structured light consistency information and Helmholtz normal information. As will be shown later in Section 4.4, these complementary cues can efficiently be combined in a variational framework based on a continuous min-cut formulation to significantly improve the reconstruction accuracy in comparison to purely triangulation-based techniques (see Section 4.6).

4.3 Acquisition Setup

Our measurement device is designed to allow the acquisition of images from varying view-light-conditions where either structured light patterns are projected or LED light sources are used to illuminate the object. Furthermore, the setup explicitly allows the acquisition of reciprocal image pairs. Setups with similar functionality have been addressed in several publications (e.g. [HLZ10, SWRK11]) for the acquisition of 3D surfaces with reflectance information. In these works, however, the Helmholtz principle was not utilized.

In our setup, depicted in Figure 4.1, 11 cameras (SVS Vistek SVCam CF 4022COGE industrial video cameras with a resolution of 2048×2048 pixels) are mounted on a vertical arc and oriented towards a turntable. In order to acquire positional information of the object shape, 4 LG HS200G projectors (LED-DLP, 200lm) with a resolution of 800×600 pixels are installed in the vicinity of the cameras. During the measurement process, these are used to project Gray code patterns onto the object surface. 198 LED lights are additionally distributed over a hemispherical gantry. Care has been taken to place these LEDs in a symmetrical manner with respect to the turntable axis and the cameras, so that we can obtain reciprocal image pairs by rotating the turntable (as illustrated in Figure 4.1).

The geometric calibration of cameras and projectors (pinhole camera model including distortion) is obtained via a structured light based bundle adjustment (see Chapter 3). In order to estimate the turntable axis, a reference calibration target with easy-to-detect features is placed onto the turntable and the features are tracked. In addition, light source positions are estimated using four mirroring spheres which can be rotated by the turntable. Highlights on the spheres are detected under illumination by the individual LEDs and the rays which are reflected from the spheres are triangulated. The setup is also fully radiometrically calibrated via standard reflectors for the cameras and LED light sources in order to be able to exploit the Helmholtz reciprocity. This requires that the image pair measurements are conducted in high dynamic range (HDR). In addition, we also use HDR imaging for the structured light to facilitate robust pattern decoding (see Chapter 3).

With this setup, we acquire a number of images of an object. For each view- and rotation-configuration, the data consists of radiometrically corrected color values under several illuminations and the per-pixel codewords decoded from the structured light projections.

4.4 Variational Formulation

We formulate the reconstruction of the object surface δV as a variational problem. The formulation depends on a vector field of Helmholtz normals $\mathbf{H}$ and three scalar fields defined on the continuous volume $\mathbb{R}^3$: the consistency measure c, the outside count o and the visibility count v. We use $c(\mathbf{x})$ to denote the number of camera-projector consistent codewords detected at $\mathbf{x} \in \mathbb{R}^3$, $o(\mathbf{x})$ for the number of cases in which structured light triangulation has determined $\mathbf{x}$ to be outside of the object and $v(\mathbf{x})$ for the total number of configurations in which $\mathbf{x}$ was visible from the camera. From these scalar fields, we derive visibility-normalized versions $\hat{c} = \frac{c}{v}$ and $\hat{o} = \frac{o}{v}$. The resulting consistency-weighted vector field $c\mathbf{H}$ has a large magnitude in the vicinity of the surface, is aligned perpendicularly to the surface, and diminishes in a greater distance. To find the object interior $V \subset \mathbb{R}^3$, we seek to optimize the following problem

$$\min_V E(V) = \underbrace{-\lambda_1 \int_{\delta V} \langle c\mathbf{H}, \mathbf{n} \rangle \, \mathrm{d}\, A}_{E_1} + \underbrace{\lambda_2 \int_V \hat{o} \, \mathrm{d}\, V}_{E_2} + \underbrace{\lambda_3 \int_{\delta V} (\alpha - \hat{c}) \, \mathrm{d}\, A}_{E_3}, \quad (4.1)$$

where λ_1, λ_2 and λ_3 are relative weights of the individual terms and $\alpha > 1$ denotes a constant determining the minimum regularization strength within a completely consistent region. The first term E_1 considers the flux of the vector field $c\mathbf{H}$ through the object surface. This term is minimized by a surface that is perpendicular to the reconstructed Helmholtz normals $\mathbf{H}$, i.e. for the same orientation of the local surface normals $\mathbf{n}$ and the Helmholtz normals $\mathbf{H}$, and in regions with a high consistency c. The second term E_2 is used as an outside constraint to penalize regions of large values $\hat{o}$. This prevents the algorithm from short-cutting through concavities. The last term E_3 represents a regularization term and enforces a

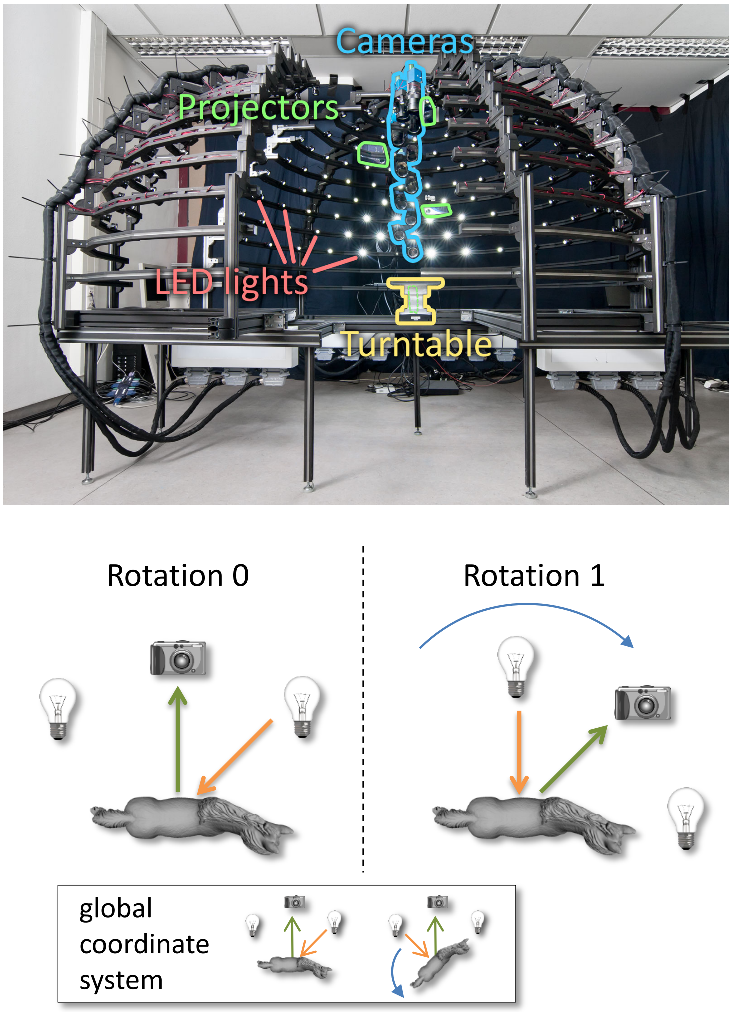


Figure 4.1: *The reciprocal measurement setup. Top: one quarter of the light-dome is slid open to provide a better view on the inside. Bottom: Illustration of symmetrical light setup. The images at the bottom illustrate the actual rotation of the object on the turntable in the measurement setup. The upper images are shown in the object coordinate system and demonstrate how this rotation leads to reciprocal image pairs.*

minimal surface. This penalty is weighted with the consistency $\hat{c}$ obtained from the structured light.

Both the structured light consistency and the technique used to estimate Helmholtz normals will be described in the following subsections.

4.4.1 Structured Light Consistency

For each point $\mathbf{x} \in \mathbb{R}^3$ and all combinations (i, j, k) of rotations, cameras and projectors, we perform an independent classification. Each time, we determine whether the point is consistent, lies before or behind the surface seen in the respective pixel in the camera or cannot be reliably classified at all (see Figure 4.2). We consider a point as consistent if the codeword decoded in the camera image j agrees with the codeword cast by the projector k. This can easily be checked by projecting the point into both the camera and the projector and comparing the codewords. All points meeting this criterion form the consistency volume depicted in green hatching in Figure 4.2. When the point does not lie within this region, we classify it according to its position along the ray r_c from the camera. It can either lie in front of the intersection of r_c with the surface $\mathbf{x}_s$, and thus outside of the object, or behind the intersection, in which case this sample does not provide any information at all. In practice, we distinguish these cases by triangulating the position $\mathbf{x}_s$. This is possible since we can determine the projector ray r_p from the codeword which lies at the projection of $\mathbf{x}$ into the camera image. In order to incorporate the inaccuracy of triangulation, we consider points that are not consistent but still lie closer to $\mathbf{x}_s$ than a user-selected threshold δ as unreliable. In case the triangulation fails completely, since r_c and r_p do not intersect or no codeword was found in the camera image, the sample combination is discarded for $\mathbf{x}$.

After performing this classification for every combination (i, j, k), we count the number of consistent combinations $c(\mathbf{x})$, combinations in which the point was determined to lie in front of the surface $o(\mathbf{x})$, and the number of unreliable combinations $u(\mathbf{x})$. We then set the visibility to the number of samples that were not discarded ($v = c + o + u$) and derive the normalized consistency and outside probabilities $\hat{c}$ and $\hat{o}$.

4.4.2 Helmholtz Normal Estimation

At each point, the corresponding normal information is estimated via the Helmholtz principle [ZBK02]. This principle is based on the reciprocity of the BRDF, i.e. $\rho(\mathbf{v}, \mathbf{l}) = \rho(\mathbf{l}, \mathbf{v})$, and can be utilized wherever several image pairs are available in

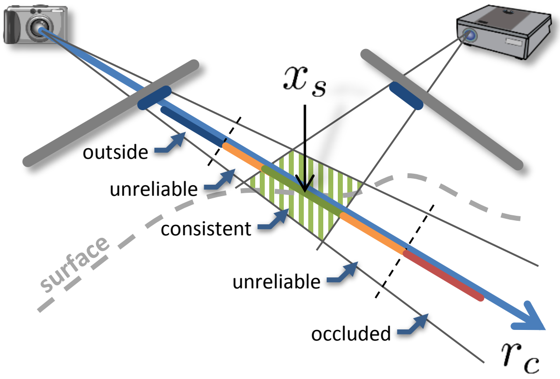


Figure 4.2: *Structured light consistency terminology illustration: The cones through the decoded codeword in the camera image and the known source-pixel in the projector plane intersect to form a consistency volume. The object surface intersects this volume. All positions* $\mathbf{x}$ *along the viewing ray* r_c *are classified as consistent, unreliable, being in front of or behind the surface.*

which the position of the light source and the camera have been exactly exchanged. Instead of actually moving the light source and camera, these image pairs can also be acquired by rotating the object and using two light sources, mounted symmetrically around the camera (see Figure 4.1). Unfortunately, accurately achieving this symmetry in a setup is difficult for practical reasons. For example, the emitter of the light source and the projective center of the camera are difficult to align. In a larger setup, there are also physical restrictions for the available mounting positions due to the size of the components and the used gantry configuration. Together, these restrictions might result in errors of a few degrees.

For this reason, we relax the assumption of perfect Helmholtz stereopsis and instead assume that the BRDF is locally smooth enough to allow for barycentric interpolation between three light samples. Therefore, highly specular materials cannot be handled with this approach and would require the development of alternative schemes. We first compute a spherical Delaunay triangulation of the set of available light source directions. Given an (idealized) pair of reciprocal camera/light-positions $\mathbf{y}$ and $\mathbf{y}'$, we obtain the two corresponding directions for each point $\mathbf{x}$ as $\mathbf{d}$ and $\mathbf{d}'$ and find the respective spherical triangles. These provide us with barycentric interpolation weights α_i, α_i' and light directions $\mathbf{l}_i, \mathbf{l}_i'$ with $i \in \{1, 2, 3\}$ in such a way that

$$\rho(\mathbf{d}, \mathbf{d}') \approx \sum_{i=1}^{3} \alpha_i' \rho(\mathbf{d}, \mathbf{l}_i') \tag{4.2}$$

and

$$\rho(\mathbf{d}', \mathbf{d}) \approx \sum_{i=1}^{3} \alpha_i \rho(\mathbf{d}', \mathbf{l}_i). \tag{4.3}$$

For each of these BRDF samples, we now have actual radiance measurements I_i and I'_i available. We assume, that these measurements have already been radiometrically corrected to compensate for the light fall-off at the point $\mathbf{x}$ by using the calibration of the measurement setup. Hence, we have

$$\rho(\mathbf{d}, \mathbf{l}'_i) = \frac{I_i}{\langle \mathbf{n}, \mathbf{l}'_i \rangle} \tag{4.4}$$

and

$$\rho(\mathbf{d}', \mathbf{l}_i) = \frac{I'_i}{\langle \mathbf{n}, \mathbf{l}_i \rangle}. \tag{4.5}$$

Since we assume from the Helmholtz reciprocity that $\rho(\mathbf{d}, \mathbf{d}') = \rho(\mathbf{d}', \mathbf{d})$, we can estimate the normal by solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{n}} E_n(\mathbf{n}) &= \frac{1}{\sum_{j=1}^{N} w^{(j)}(\mathbf{n})} \\ &\quad \cdot \sum_{j=1}^{N} w^{(j)}(\mathbf{n}) \left\| \sum_{i=1}^{3} \alpha_i'^{(j)} \frac{I_i^{(j)}}{\langle \mathbf{n}, \mathbf{l}_i'^{(j)} \rangle} - \sum_{i=1}^{3} \alpha_i^{(j)} \frac{I_i'^{(j)}}{\langle \mathbf{n}, \mathbf{l}_i^{(j)} \rangle} \right\|^2 \end{aligned} \tag{4.6}$$

The weight terms

$$w^{(j)}(\mathbf{n}) = (\langle \mathbf{n}, \mathbf{d}^{(j)} \rangle \langle \mathbf{n}, \mathbf{d}'^{(j)} \rangle)^2 \tag{4.7}$$

are included to reduce the relative importance of samples at grazing angles as these measurements are less accurate and the division through the small cosine term would further increase these errors. If either of the two scalar products becomes smaller than a threshold, we set $w^{(j)}(\mathbf{n}) = 0$. Only samples which were classified as consistent by the structured light computation are included in the estimation to avoid the influence of occlusion and shadowing. We solve the resulting non-linear optimization problem with a Levenberg-Marquardt optimizer [Lev44] using the SVD approach proposed in [ZBK02] as initialization.

4.5 Implementation Details

Storing all values needed for the reconstruction in a regular grid would be prohibitively expensive with regard to memory consumption and computational demands, preventing reconstructions at high levels of detail. Therefore, we employ an octree-based data structure that adapts the resolution to the shape of the object. For this, we use an iterative algorithm which successively refines the octree based on the previous reconstruction.

Structured Light Based Octree Initialization: Starting with a tree at an initial coarse resolution, we successively refine cells as long as a common structured light codeword appears in a sufficiently large number of the re-projected footprints of the cell. When the surface runs through the cell, the same codeword should lie within the footprints in a large number of the images. In contrast, the number of coincidentally identical codewords decreases with the size of the cell. The refinement continues up to a maximum initial octree depth.

Octree Update: The consistency values $\hat{o}$ and $\hat{c}$ and the Helmholtz normals $\mathbf{H}$ are all evaluated and stored at the corners of each octree cell. Consequently, the computation of these values is only necessary for the corners newly added by the subdivision process. All previously existing corners in the tree remain unchanged and hence do not require being updated.

Continuous Min-Cut: Having the consistency values and the normals in the volume of interest, we seek to find a globally optimal min-cut that partitions the volume into inside and outside. Using the Gauss-Ostrogradsky theorem

$$\int_{\delta V} \langle c\mathbf{H}, \mathbf{n} \rangle \;\mathrm{d}\,A = \int_V \operatorname{div}(c\mathbf{H}) \;\mathrm{d}\,V, \tag{4.8}$$

we can represent the term E_1 in Functional (4.1) as an integral over the volume. The total energy can then be directly mapped onto the continuous min-cut functional

$$\min_\lambda D(\lambda) = \int_\Omega (1-\lambda)C_s + \lambda C_t + C|\nabla\lambda| \;\mathrm{d}\,x \tag{4.9}$$

given in [YBT10]. For this, we set

$$\begin{aligned} C &= \hat{c} && (4.10)\\ C_s &= \max(0, \operatorname{div}(c\mathbf{H}) - \hat{o}) && (4.11)\\ C_t &= \max(0, -\operatorname{div}(c\mathbf{H}) + \hat{o}). && (4.12) \end{aligned}$$

To obtain the desired labeling λ of the volume, we are able to apply the continuous max-flow algorithm proposed in [YBT10] due to the duality between max-flow and min-cut.

In contrast to [YBT10], in which a dense 3D grid was employed, we perform the max-flow computation directly on the proposed octree structure. For this, we compute C, C_s and C_t for each of the cells. To obtain the divergence within the cell, we compute samples of $c\mathbf{H}$ for each of the cell's facets by averaging the corresponding values stored at the surrounding corners. Weights based on the Helmholtz normal error E_n (see Equation (4.6)) are employed to decrease the influence of unreliably estimated normals. These samples of the vector field are then used to approximate the flow through each of the facets and, consequently, the average divergence within the cell via the divergence theorem. We also weight these divergences with the average reliability of the Helmholtz normal estimation to avoid large divergences in regions with unreliable normals.

Once the surface has been computed, all octree cells with a distance to the reconstructed surface smaller than twice the cell's edge length are refined to adaptively increase the octree resolution near the surface. Furthermore, the maximum allowed octree level is increased in each iteration to successively refine the reconstruction.

Smooth Surface Reconstruction: Especially in regions where the divergence term dominates the min-cut reconstruction, the isosurfaces of the resulting continuous labeling are not smooth. To infer a smooth final surface, we compute a signed distance function f. We constrain this function to lie within a band of one octree cell around the reconstructed isosurface. To incorporate the normal information obtained via Helmholtz stereopsis, we search for a surface corresponding to the vector field $\mathbf{H}$ that is as smooth as possible. This can be achieved by using an energy function similar to the one proposed in [CT11]. We utilize the same normal consistency and smoothness terms. However, since we do not have a point cloud as input, we remove the data term which penalizes deviations from the points. Instead, we use hard constraints which enforce the implicit function f to be negative inside the object and positive on the outside and perform the optimization only within the narrow band. This results in the following optimization problem:

$$\begin{aligned} \min_{f} \quad & \mu_0 \sum_{i=1}^{N} w_i \|\nabla f(x_i) - \mathbf{n}_i\|^2 + \mu_1 \frac{1}{|V|} \int_V \|Hf(x)\|^2 dx, \\ \text{s.t.} \quad & \begin{array}{ll} f(x) < -\epsilon & \text{for all } x \text{ at the inner border of the band} \\ f(x) > \epsilon & \text{for all } x \text{ at the outer border of the band} \end{array} \end{aligned} \tag{4.13}$$

where the weights w_i are used to consider the normal estimation error, H represents the Hessian matrix, and V denotes the volume of the involved band. This is a positive definite quadratic optimization problem with linear inequality constraints. We employ an active set block pivoting method [Adl00] using a conjugate gradient solver with a limited memory incomplete Cholesky factorization [LM99b] as preconditioner to take advantage of the sparsity of the problem. The regularization term utilized in the SSD penalizes curvature. In contrast, the min-cut optimization employs a minimum surface regularization. It is not trivial to use the same regularization in both cases due to the fact that two different optimization techniques are used. However, as the signed distance function is restricted to a narrow band around the min-cut solution, a strong deviation of the smooth signed distance result from the min-cut solution is prevented. The final step in our reconstruction consists of converting the refined implicit function into a mesh. This is achieved by applying the isosurface extraction on octree grids presented in [KKDH07].

4.6 Experimental Results

We evaluate our technique both on a synthetic reference dataset as well as real world examples. All test cases shown, including the synthetic ones, were created for $15°$ steps of the turntable rotation and contain for each rotation a full measurement from our setup with 11 cameras, 4 projectors, and 198 LED light sources.

In Figure 4.3, we show our results on the synthetic test dataset. For this experiment, we rendered images under a glossy reciprocal Phong BRDF and simulated projector-codeword images of the Lucy mesh [TSDSR]. A comparison with a purely structured light based reconstruction exploiting projector super-resolution [WSRK11] is shown. This technique first reconstructs a point cloud and then utilizes Poisson surface reconstruction [KBH06] to obtain a mesh. The reconstruction obtained with the technique introduced in the scope of this chapter remains below a deviation of one octree cell from the true surface for almost the complete surface. In comparison, the reconstruction based on the technique presented in Chapter 3 shows larger errors in the concave regions of the figurine that are only visible in a few cameras and not accurately reconstructed via the Poisson reconstruction.

In Figure 4.4, we show a reconstruction of a real-world brass figurine. This demonstrates that our approach presented in this chapter can cope with glossy materials and reconstruct even fine surface details, which are lost in the reconstruction obtained with a high-precision line laser scanner and the structured light reconstruction following our technique presented in Chapter 3 due to noise in the triangulated points. Figure 4.5 illustrates that even very small details far below the resolution of the projector patterns such as the engraved text can be reconstructed.

All reconstructions were performed on a level 10 octree (1024^3 nodes on the finest level). Using a parallelized C++ implementation, the computation of the Lucy dataset took about 8 hours on a computer with two Intel Xeon E5620 processors and 24GB of RAM. However, depending on the octree occupancy and the selected smoothing strength, timings might vary from case to case.

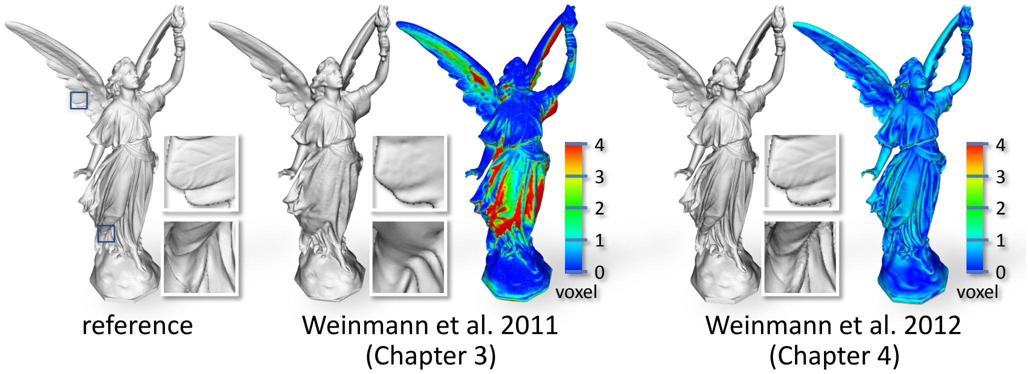


Figure 4.3: *Comparison on synthetically generated input data: The technique presented in this chapter preserves fine details, e.g. on the wings or the garment, and also reconstructs the concave regions, which are otherwise lost during the Poisson reconstruction step in the technique presented in Chapter 3. The plots show the Hausdorff distance to the reference mesh, normalized to the size of an octree cell. Note that our reconstruction remains within an error of about one octree cell in almost all places.*

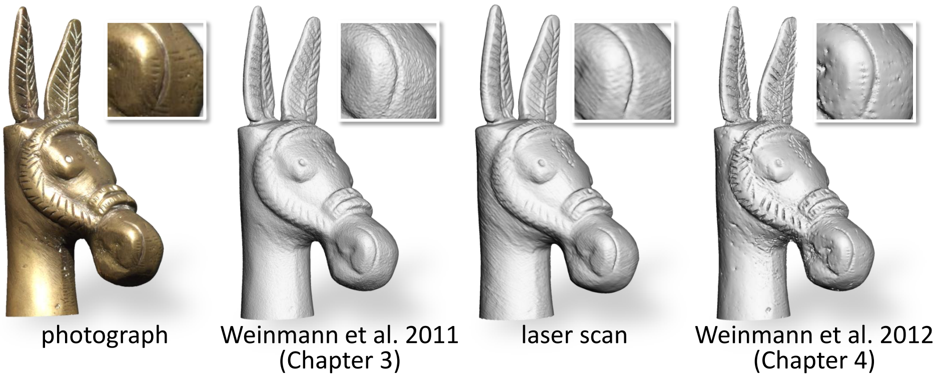


Figure 4.4: *Comparison of results on a glossy brass figurine: The laser scan was created with a high-precision line laser scanner mounted on a measuring arm with a total accuracy of about* $60\mu m$*. The reconstruction using the technique presented in the scope of this chapter shows considerably more fine surface details in comparison to both the reconstruction via laser scanning and the reconstruction using our triangulation-based technique presented in Chapter 3.*

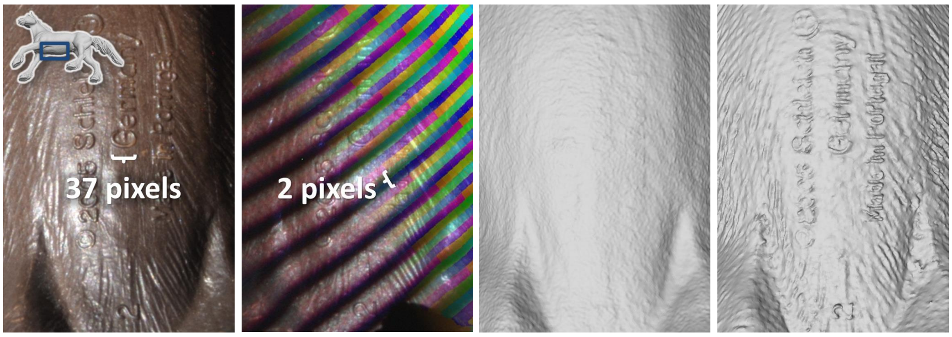


Figure 4.5: *A detail view of a toy horse. The first image is taken from the input data of the Helmholtz normal estimation and illustrates the resolution of the cameras. The second image shows one of the structured light images overlaid with the decoded codewords and demonstrates the resolution of the projector. The last two pictures compare the reconstruction results obtained using the method proposed in Chapter 3 (published in [WSRK11]) (left) and the method described in this chapter (published in [WRO+12]) (right). The reconstructed writing exhibits a depth of approximately* $150\mu m$*.*

4.7 Conclusions

The method presented in the scope of this chapter is based on the combination of structured light scanning with Helmholtz normal estimation and has been demonstrated to allow for the reconstruction of high-quality 3D models, faithfully representing even fine surface details. By applying a variational approach, we formulate the 3D reconstruction as one combined optimization problem over all available input data. In particular, the additionally used normal information represents a crucial prerequisite to overcome the problem of triangulation-based approaches which suffer from noise induced by e.g. inaccurately localized correspondences and inaccuracies in calibration and are therefore limited with respect to the reconstruction of fine surface structures. However, the use of Helmholtz normals limits the range of materials that can reliably be reconstructed to the class of opaque materials in comparison to our triangulation-based multi-camera, multi-projector super-resolution framework for structured light proposed in Chapter 3. The latter method allows an accurate reconstruction for the wider range of diffuse to even highly specular materials as long as a sufficient diffuse component is present in the reflectance behavior.

Particularly for very specular materials, the results obtained with the method described in this chapter could be further improved by adding outlier robust statistics to the Helmholtz normal estimation. For very challenging cases, these could also be complemented with alternative normal estimation techniques such as shape-from-specularity techniques, as e.g. investigated in the following chapter, to handle an even wider range of materials. All the normals obtained by applying the different methods could e.g. be combined per octree corner and, depending on a certain consistency measure, the most appropriate normals with respect to the respective surface material might be selected. We will further discuss this aspect in Part IV. Further improvements might be achieved when integrating additional consistency terms into the optimization, such as color consistency.

Finally, the possibility of performing reconstructions for large objects at high resolutions is currently limited by the time needed for the graph-cut and normal computations. Faster computations could probably be achieved using GPU implementations. However, this will require techniques to split the problem into smaller meaningful subproblems, due to the limited memory of GPUs.

CHAPTER 5

3D RECONSTRUCTION OF MIRRORING OBJECTS

In the scope of this chapter, we complement our investigations in Chapter 3 and Chapter 4 by focusing on 3D shape acquisition for mirroring objects. In particular, we present a novel, robust multi-view normal field integration technique that allows the reconstruction of the full 3D shape of mirroring objects. We employ a turntable-based setup with several cameras and displays. The latter ones are used to display illumination patterns which are reflected by the object surface. The pattern information observed in the cameras allows the calculation of individual volumetric normal fields for each combination of camera, display and turntable angle. As the pattern information might be blurred depending on the surface curvature or due to non-perfect mirroring surface characteristics, we locally adapt the decoding to the finest still resolvable pattern resolution. In complex real-world scenarios, the normal fields contain regions without observations due to occlusions and outliers due to interreflections and noise. Therefore, a robust reconstruction using only normal information is challenging. Via a non-parametric clustering of normal hypotheses derived for each point in the scene, we obtain both the most likely local surface normal and a local surface consistency estimate. This information is utilized in an iterative min-cut based variational approach to reconstruct the surface geometry. The developed technique allows highly accurate 3D reconstructions of mirroring objects both for synthetic data and real-world data. The method presented in this chapter has successfully undergone peer review (see [WORK13]) and, to the best of our knowledge, currently still represents the state-of-the-art technique regarding the reconstruction of the full 3D shape of mirroring objects with a more complex surface geometry.

After a discussion of the limitations of previously published techniques and an overview of the contributions presented in this chapter (see Section 5.1), we formulate the task of reconstructing mirroring objects in terms of a variational framework (see Section 5.2). This is followed by a description of our acquisition setup (see Section 5.3) and our acquisition approach (see Section 5.4). Finally, we evaluate our method based on several experiments (see Section 5.5) and provide respective conclusions (see Section 5.6).

5.1 Motivation

While a huge number of 3D acquisition techniques has been developed so far, today's challenges can be found when considering surfaces which exhibit a complex surface reflectance behavior such as mirroring objects as considered in this chapter. For such objects, most traditional techniques, such as laser scanners, structured light or multi-view stereo, are not applicable. Instead, the problem of reconstructing mirroring surfaces has been approached in various different ways as already discussed in detail in Section 2.3.3.

Under the assumption of a perfectly mirroring surface, the appearance of a surface point only depends on the surrounding environment, the viewing angle and the local surface normal. By controlling the environment, it is directly possible to estimate normal information (see e.g. [SWN88, TLGS05, CGS06, FCM$^+$08]). An alternative is to rotate the object and to track the optical flow (e.g. [RB06, AVBSZ07, SVTA10]).

Several approaches such as the ones in [CGS06, FCM$^+$08] use these normals to perform a single-view normal field integration and are thus limited to partial 2.5D reconstructions. Other approaches rely on deriving a normal consistency measure and performing a multi-view reconstruction (e.g. [BS03, NWR08]). However, normal consistency alone is not suitable to reconstruct fine surface details. Therefore, a final refinement step is performed in [NWR08] to combine the geometry estimated from the normal consistency with the observed surface normals. However, none of the mentioned approaches has shown high-quality reconstructions for complex geometries in the presence of occlusions and interreflections.

To address this problem, we exploit the fact that outliers due to occlusions or interreflections are not consistent for different measurements taken under varying configurations of viewpoint and light source position. Inspired by the multi-view normal field integration approach presented in [CLL07] but utilizing a numerical scheme to obtain a globally consistent surface reconstruction similar to our approach described in Chapter 4, we formulate the problem in terms of an optimization which combines both a local surface consistency measure and the observed normal information. We determine these quantities in an outlier-robust way via mean-shift clustering [Che95] of the individual local normal hypotheses which result from different configurations of viewpoint and illumination. This makes our approach capable of handling occlusions. To acquire the full shape of the considered object, the utilized setup comprises a turntable in combination with eleven cameras and three screens that display structured light patterns which are reflected by the object surface. Our technique produces high-quality reconstructions of the full 3D shape of an object not only on synthetic but also on real-world data (see Figure 5.1). In contrast to the previously presented multi-view normal field integration approaches [CLL07, Dai09], which only allow reliable reconstructions for synthetic data, our method only requires normal information and does not rely on using additional silhouette information, which is difficult to determine for specular objects.

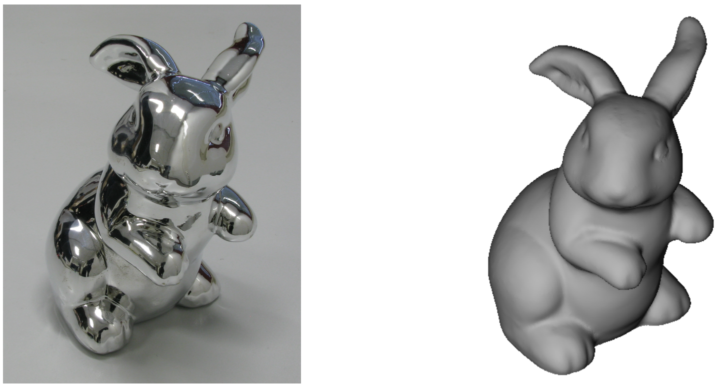


Figure 5.1: *Bunny figurine and reconstructed model.*

In summary, the key contributions of our technique presented in the scope of this chapter include:

- a system that allows the acquisition of the full 3D shape of mirroring objects based on multi-view normal field integration, and
- a novel clustering-based scheme to integrate different volumetric normal fields which is robust in the presence of outliers and noise and makes accurate 3D reconstructions possible on real-world data.

5.2 Problem Statement

Given a set of κ_c calibrated cameras which are positioned to observe a mirroring object from different viewpoints and a set of κ_s screens, our goal is to reconstruct the object surface δV of an object with the volume V by utilizing only normal information recovered for the individual views. Apart from a smoothness prior, we do not incorporate any prior knowledge about the object geometry such as the assumption of rather flat surfaces [CGS06, FCM$^+$08] or an initial visual hull reconstruction [CLL07]. Furthermore, our approach should consider the possibility of self-occlusions of the object geometry. Due to the complexity of real-world scenarios, we also have to design our reconstruction technique to be robust to noise. In addition, violations of the assumed reflectance model need to be handled to some degree as well as incomplete normal fields, which occur when no normal information can be derived for certain parts of the object surface. For this reason, we formulate the surface reconstruction as a variational energy minimization problem similar to [CLL07] according to

$$\min_{V} \left\{ -\lambda_1 \int_{\delta V} \langle c\mathbf{N}, \mathbf{n} \rangle \, \mathrm{d}A + \lambda_2 \int_{\delta V} \alpha \, \mathrm{d}A \right\}, \tag{5.1}$$

where λ_1 and λ_2 denote weighting coefficients, c represents a scalar field of surface consistency and the consistency-scaled vector field $c\mathbf{N}$ contains information about both the local probability of surface presence and the local normal information for the points in the volume and α represents a regularization parameter. The first term in Functional (5.1) is minimized for high consistency values and a surface which is perpendicular to the observed normals $\mathbf{N}$. The local normals of the estimated surface are denoted with $\mathbf{n}$. The second part of the functional represents a regularization term which enforces a minimal surface area to avoid overfitting by increasing the cost for oscillating surfaces. Similar to our technique presented in Chapter 4, the global optimization of this functional can be mapped to the optimization of the continuous min-cut functional [YBT10]

$$\min_{\lambda} \left\{ \int_{\Omega} (1-\lambda) C_s + \lambda C_t + C \left| \nabla \lambda \right| \mathrm{d}V \right\} \tag{5.2}$$

via specifying

$$\begin{aligned} C &= \lambda_2 \, \alpha && (5.3) \\ C_s &= \lambda_1 \max \{0, \operatorname{div}(c\mathbf{N})\} && (5.4) \\ C_t &= \lambda_1 \max \{0, -\operatorname{div}(c\mathbf{N})\} \, . && (5.5) \end{aligned}$$

We choose this formulation as it provides efficiency concerning memory consumption and alleviates metrification errors.

After describing the utilized setup in the following section, we describe the technique to acquire and integrate the normal information (see Section 5.4).

5.3 Acquisition System and Calibration

For the acquisition, we use a turntable-based setup as illustrated in Figure 5.2, where eleven cameras with a resolution of 2,048 × 2,048 pixels are positioned on

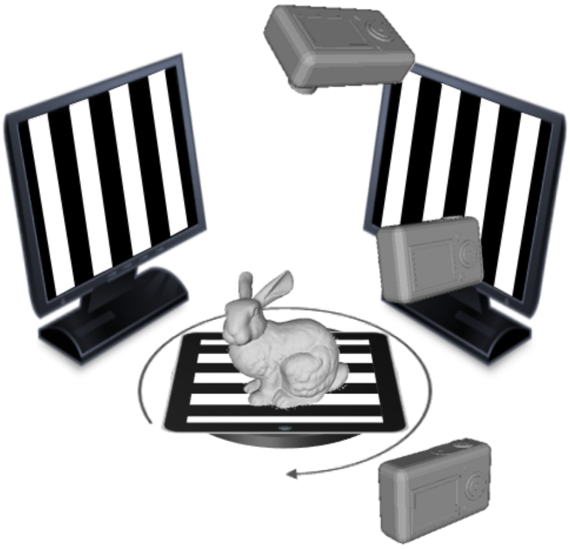


Figure 5.2: *Sketch of the utilized setup: The screens successively display the series of patterns for each rotation of the turntable. The reflected pattern on the object surface is observed by the cameras. For illustration purposes, only three of the eleven employed cameras are drawn.*

a vertical arc. The calibration of the cameras and the turntable axis is performed by using a rotating three-dimensional calibration target with robustly detectable markers as already used for the calibration of the setup used in Chapter 4 and explained in detail in [SSW$^+$14]. Similar to e.g. [FCM$^+$08, NWR08], we use a monitor-based shape-from-specularity approach to simulate a dense illumination area. Two static displays with resolutions of $2{,}048 \times 1{,}152$ pixels and $2{,}560 \times 1{,}600$ pixels are placed close to the objects to display patterns. Gray code patterns and their inverses are used for the unique identification of the reflection of each screen pixel on the mirroring surface with a small number of acquired images. In order to illuminate the object surface as completely as possible, we place the object onto the display of an Asus TF300T-1E031A tablet with a resolution of $1{,}280 \times 800$ pixels, which is on top of the turntable and also used to display patterns. Both the monitor displays and the tablet display need to be chosen to provide a good coverage of the sphere of possible reflection directions. Additionally, we identified that it is important that the tablet is stable enough to support the object weight when placing the object on it, i.e. tablets with hard glass surfaces are more suitable. In turn, this results in interreflections which have to be taken into account during the reconstruction.

To calibrate the screen position and the position of the tablet display, we use the decoded pattern information observed in the images of the involved cameras and perform an estimation of the display pixel positions $\mathbf{x}_l$ via triangulation so that the resulting point cloud represents (a part of) the display. From the decoded bits for each of the m points in the point cloud, it is possible to uniquely determine its offset $\mathbf{u}_l = [u_l, v_l]^T$ from the origin $\mathbf{o}$ of the display frame which we consider to be at the upper-left. Using this information, we can derive the coordinate frame of the screen consisting of the origin $\mathbf{o}$ and the spanning vectors $\mathbf{a}$ (parallel to the display width) and $\mathbf{b}$ (parallel to the display height) via optimizing

$$Q = \sum_{l=1}^{m} (\mathbf{x}_l - (\mathbf{o} + u_l\,\mathbf{a} + v_l\,\mathbf{b}))^2. \tag{5.6}$$

The resulting linear system is solved using least squares minimization. Thus, given the screen calibration, we can directly determine the 3D location of a pixel on the screen by considering its bit sequence.

For the calibration of the screens, it is not necessarily required to see the complete screens in the camera images as several parts of the displays seen in different cameras are sufficient. While our calibration method requires the monitor to be close to the object, this is eventually desirable for the measurement to cover a larger part of the mirroring surface with the projected patterns and reduce the influence of light fall-off.

5.4 Multi-View Shape-from-Coded-Illumination

In order to bring classical shape-from-specularity techniques to the multi-view scenario, we first discuss the utilized encoding of the illumination patterns as well as the problems occurring due to surface curvatures which we solve via a fuzzy decoding of the patterns. Subsequently, we describe how the decoded information is used to generate normal hypotheses from which the normal field required in the optimization, i.e. Functional (5.1), and the surface consistency are derived. An overview of our method can be found in Algorithm 1 and the respective block diagram is shown in Figure 5.3.

5.4.1 Coded Illumination

In order to encode the illuminations coming from the displays, we use Gray code patterns which make a robust decoding possible. Additionally, similar to the approach presented in [Tro95], we take the inverse patterns to increase the

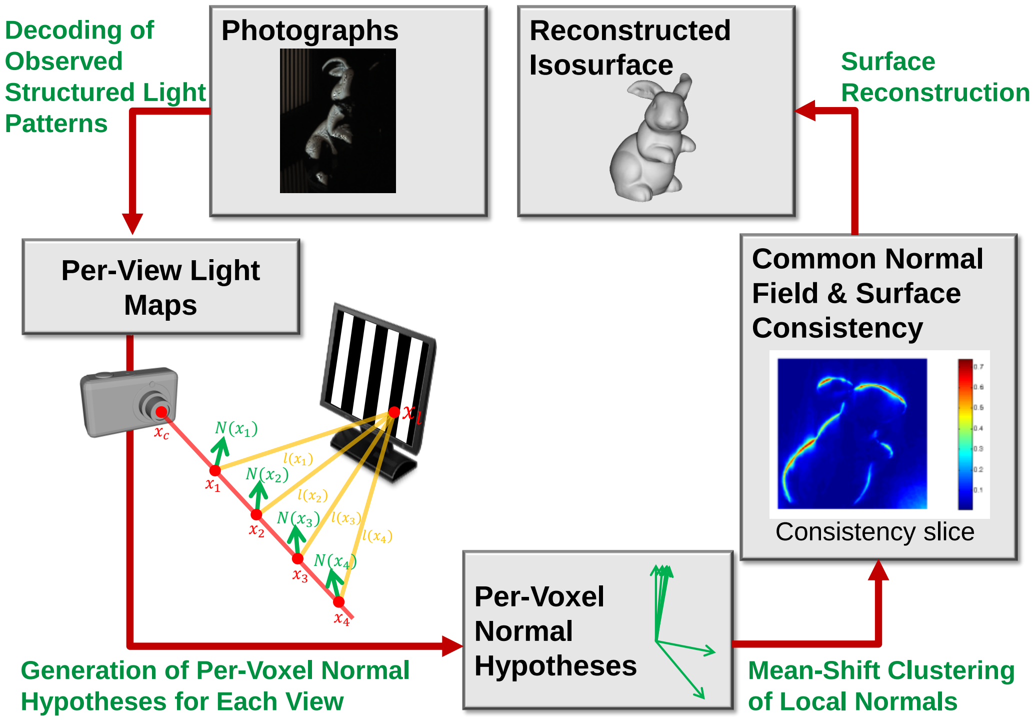


Figure 5.3: *Block diagram of the proposed method.*

Algorithm 1 Proposed Multi-View Normal Field Integration Approach for 3D Reconstruction of Mirroring Objects.

```
projection of structured light patterns onto the mirroring surface (for each camera/screen/rotation configuration)
computation of light maps by decoding the structured light sequences (for each camera/screen/rotation configuration)
for all voxel corners do
    get structured light decodings from each of the light maps
    calculate one normal hypothesis for each of the light maps
    obtain surface consistency and common normal field entry using mean-shift clustering of all normals
end for
perform surface reconstruction similar to our technique described in Chapter 4
```

robustness. To decode the displayed bit sequences, we compare the intensity values observed at each pixel $\mathbf{u}$ in the pair consisting of image $\mathcal{I}_{i,j,k,q}$ seen while displaying pattern $\mathcal{P}_q$ and image $\bar{\mathcal{I}}_{i,j,k,q}$ seen while displaying its inverse pattern $\bar{\mathcal{P}}_q$. The parameter $i = 1, \ldots, \kappa_c$ denotes the camera index, the parameter $j = 1, \ldots, \kappa_s$ denotes the screen index, the parameter $k = 1, \ldots, \kappa_r$ denotes the rotation index, and q denotes the index of the pattern. If the difference is below a certain threshold, we mark the decoded bit as unreliable. We use

$$\left|\mathcal{I}_{i,j,k,q} - \bar{\mathcal{I}}_{i,j,k,q}\right| < 0.1\,\mathcal{I}_{i,j,k,0}, \tag{5.7}$$

where $\mathcal{I}_{i,j,k,0}$ represents the photo taken under illumination by the fully lit pattern.

As each pixel on the displays can be uniquely encoded and its 3D position on the screen is known from the screen calibration, observed codewords can directly be related to the corresponding 3D positions on the screen. Hence, we generate a light map [BW10, CGS06] for each individual camera under each rotation angle k of the turntable and under illumination from each display j. These light maps $\mathcal{L}_{i,j,k}$ assign the respective light source position to each pixel in the camera image. In general, there will not be observations for all the pixels. The reason for this is that, depending on the shape of the object and the position of the illuminant, only a part of the surface will reflect patterns towards the camera.

Interreflections introduce outliers in the light maps. In addition, depending on the curvature of the mirroring surface and the different relative distances to the display pixels or other effects, such as non-ideal or spatially varying reflectance properties, it is usually not possible to decode the complete bit sequence correctly. High-frequency patterns might appear blurred on the object surface which has already been observed in e.g. [FCM$^+$08, FCMB09, BW10], and it is not possible to decide if pattern $\mathcal{P}_q$ or its inverse $\bar{\mathcal{P}}_q$ has been displayed. As a consequence, we introduce a fuzzy decoding. The basic idea is to only use the reliably decoded bits per pixel to identify the corresponding display area which illuminated this pixel. If less bits can be reliably decoded, the ambiguity in the region of the display which illuminated the pixel increases. The corresponding light source position is determined as the center of this reliably decoded region.

To address noisy decodings in the light maps, which could represent problems for the calculation of normals and, hence, also for the normal field integration algorithm, we additionally perform a subsequent filtering step. In this step, all decoded labels with less than t_{bits} reliably decoded bits for both horizontal and vertical stripe patterns are discarded. In order to calibrate the screens, we use $t_{\text{bits}} = 9$ as a very accurate decoding is possible. During the reconstruction, we use $t_{\text{bits}} = 5$. Furthermore, we also consider the average of the individual contrasts observed for the individual patterns and their inverses for each image pixel per series of patterns to filter out unreliable decodings. In principle, the quality of the decodings can be used as weights for the quality of the normals derived from them which might be included in future work.

5.4.2 Generation of Normal Hypotheses

The light maps described in the previous subsection are used to derive information about surface normals. As our setup violates the assumption of distant illumination and the object surface is unknown a-priori, the ambiguity concerning the depth of the surface along the view directions for the individual cameras cannot be neglected as in the case of far-field illumination. In our variational formulation, we therefore consider a volumetric representation to resolve this problem. In particular, the normal hypotheses are calculated separately for all the points along the view direction per pixel in each camera similar to [BS03] by utilizing the information stored in the light maps. For each point $\mathbf{x}$ in the volume and each combination of camera index $i = 1, \ldots, \kappa_c$, screen index $j = 1, \ldots, \kappa_s$ and rotation index $k = 1, \ldots, \kappa_r$, we compute a normal estimate $\mathbf{n}_{i,j,k}(\mathbf{x})$. Assuming that the object remains fixed and cameras and displays are rotated, we consider the coordinate $\mathbf{x}$ relative to the turntable. Therefore, we obtain light directions $\mathbf{l}_{j,k}(\mathbf{x})$ and view directions $\mathbf{v}_{i,k}(\mathbf{x})$ which depend on the position in the volume $\mathbf{x}$ and both on the rotation index k and the screen index j or camera index i, respectively. Following the law of reflection, we obtain the normal estimate $\mathbf{n}_{i,j,k}(\mathbf{x})$ as the bisector between $\mathbf{l}_{j,k}(\mathbf{x})$ and $\mathbf{v}_{i,k}(\mathbf{x})$. At points close to the surface, normal hypotheses derived for different camera/screen/rotation configurations, for which the corresponding points are visible, only have a small variance and almost coincide with the true surface normal. In contrast, hypotheses contradict each other at points distant to the true surface.

However, as the cameras might directly observe certain parts of the displays as well, the light maps do not only contain information about the object to be reconstructed. For the reconstruction of the object geometry, these regions in the light maps should not be propagated into the volume in the process of generating normal hypotheses. For this reason, our method also analyzes the 3D distance between the intersection of the backprojected rays with the estimated plane of the active display and the light source position stored in the light map (see Figure 5.4). If this distance is small (we use a threshold of 3mm), it is a hint that the information stored in the light map belongs to the screen geometry and can be masked out.

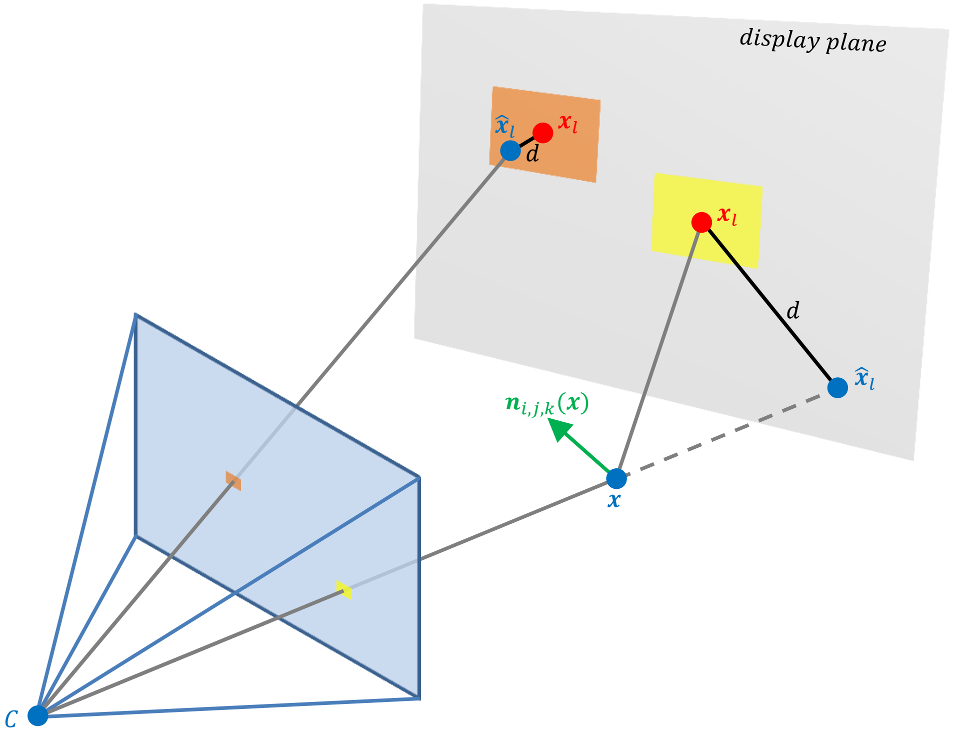


Figure 5.4: *Detection of decodings that are directly observed on the display: The image location of an observed decoding is backprojected into the volume and the intersection point $\hat{\mathbf{x}}_l$ with the plane of the active display is computed. If a structured light decoding is directly observed on the display, the distance between $\hat{\mathbf{x}}_l$ and the true light source position $\mathbf{x}_l$ is rather small (orange). In contrast, this distance is typically larger if the decodings are directly observed on a point $\mathbf{x}$ on the surface of the mirroring object between the display and the camera (yellow).*

5.4.3 Multi-View Normal Field Integration and Surface Consistency Estimation

The result of the aforementioned normal calculation step is a set of normal fields assigned to the involved camera/screen/rotation configurations (i, j, k). These individual fields need to be combined to one common normal field which contains information about the best local normal and the surface consistency.

After combining the information in the volume of interest, we have several normal votes for the different points in the considered volume. In order to find the true surface, we assume that, at a certain location $\mathbf{x}$ on the object surface, the normal hypotheses from the different cameras agree with each other and with the true surface normal. In contrast, normal estimates from the different configurations (i, j, k) will contradict each other farther away from the surface. However, due to effects such as outliers, noise, non-ideal calibration or the discretization of the volume, perfectly matching normals will hardly occur in real-world scenarios. Therefore, we can consider the observed normals as samples from an underlying probability distribution. Since the non-occluded normals should agree up to a small variance in the vicinity of the true surface, the underlying distribution should have a global maximum centered on the surface normal. Furthermore, its variance can be regarded as a measure for surface consistency. Similar measures have been used in [BS03, NWR08] for the reconstruction of highly specular and mirroring objects. As the information about the visibility of points with respect to the involved cameras is unknown, we also have to take into account that several of the normals actually come from an occluded view in addition to the noise and outliers.

Modeling the probability density of normals under occlusions is challenging as it depends on the geometry of the considered object as well as on the placement of the involved cameras and screens. Therefore, we do not model the probability density function (pdf) via a parametric model but instead only make the simplifying assumption that the density is highest for the actual surface normal. This assumption is warranted as the actual surface normal is consistent over all views where the respective surface point has been observed, whereas the outliers should not be consistent over several views. Under this assumption, finding the normal direction corresponds to finding the largest mode of the pdf. For this reason, we decided to use mean-shift clustering [Che95] as a non-parametric technique as this neither requires assuming a model nor creates discretization artifacts. We therefore define the pdf as

$$p_{\mathbf{x}}(\mathbf{n}) = \frac{1}{\kappa_c \kappa_s \kappa_r h^3} \sum_{i,j,k} K\left(\frac{\|\mathbf{n} - \mathbf{n}_{i,j,k}(\mathbf{x})\|}{h}\right), \tag{5.8}$$

and set the local normal estimate $\mathbf{N}(\mathbf{x}) = \arg\max_{\mathbf{n}} p_{\mathbf{x}}(\mathbf{n})$ to the centroid of the highest mode of the pdf. Furthermore, we use the density at the centroid as a surface consistency measure which we denote with $c(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{N}(\mathbf{x}))$.

In Equation (5.8), K represents the kernel function with bandwidth h which is chosen heuristically in the scope of our experiments. We experimented with both the Epanechnikov kernel and the Gaussian kernel and found the latter to result in a more accurate reconstruction. As an alternative, it is also possible to consider normal histograms. Then, the highest mode of the pdf corresponds to the bin with the maximum count. While this would be faster, we did not reach the quality of the reconstructions as obtained when using mean-shift clustering in the scope of our experiments.

5.4.4 Surface Reconstruction

Subsequent to calculating the estimates for the common volumetric normal field and the surface consistency as described before, we adapt the iterative optimization procedure presented in Chapter 4 to our setting. After an initialization of the utilized octree at a coarse level, the grid is successively refined according to the local surface consistency estimates in the volume. In a subsequent iterative process, the memory efficient continuous min-cut [YBT10] is applied for a global optimization per iteration. In a final step, the resulting binary indicator function is smoothed inspired by the technique presented in [CT11].

5.5 Experimental Results

We evaluate our technique in two steps. To demonstrate the robustness of our reconstruction framework, we first consider the classical multi-view normal field integration case. Here, we use per-camera normal images as input. In the next step, we show results on mirroring objects.

In a first synthetic test case, we use normal fields directly generated from the object geometry using a normal shader in OpenGL. The reconstructed model obtained using 75 viewpoints evenly distributed in the upper hemisphere is shown in Figure 5.5. Fine surface details are well preserved in the reconstruction. For all other experiments, we use 264 views ($\kappa_c = 11$ cameras are mounted on an arc and the turntable is rotated in steps of $15°$, i.e. $\kappa_r = \frac{360°}{15°} = 24$).

In order to evaluate the same scenario on real-world data, we have acquired a painted mask made of clay and estimate an independent normal map for each view using classical single-view photometric stereo [Woo80]. Subsequently, the integration is performed using our variational formulation. We use the assumption

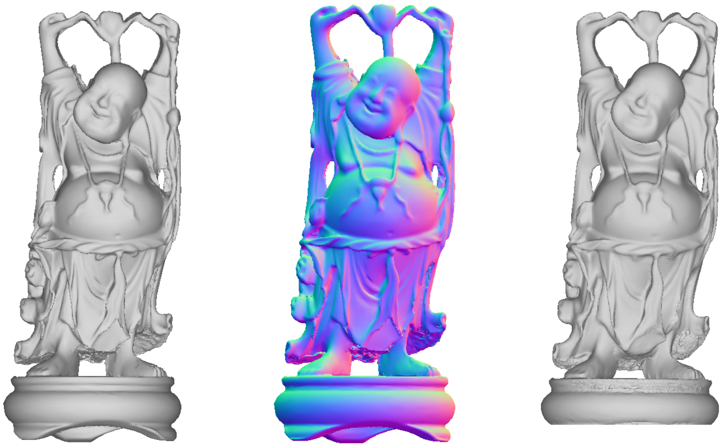
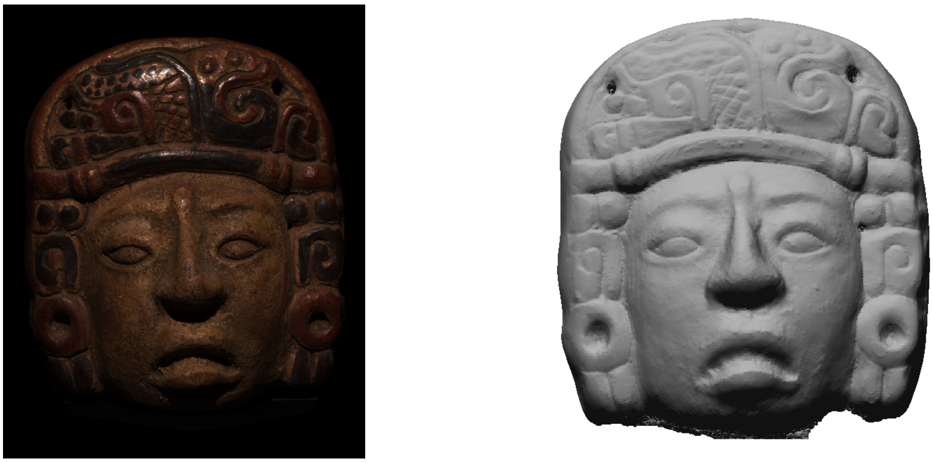


Figure 5.5: *Results on a synthetic dataset: Original Happy Buddha model [TSDSR] as seen in one viewpoint (left), observed normal field with respect to the world coordinate system (center) for this viewpoint and reconstructed model (right).*

of far-field illumination, but with our technique it would also be possible to relax this assumption by computing an individual normal at each point in the volume. As the assumption of Lambertian surface reflectance is violated due to the presence of effects such as specular reflection, shadows and interreflections on the mask surface, normal estimation based on linear least squares fitting is prone to errors. Therefore, we use a simple outlier rejection to remove the influence of too bright or too dark regions in the least squares fitting. The reconstructed model is shown in Figure 5.6. Applying a more sophisticated photometric stereo technique would probably improve the reconstruction quality.

For mirroring objects, we first consider synthetic input data for our algorithm by simulating the acquisition process. For this purpose, we represent each display via a plane which is textured according to the patterns of the Gray code sequence. The object is rotated in steps of $15°$ and observed by eleven cameras placed in the upper hemisphere. Images are then rendered for the different configurations of the involved cameras, activated screens, patterns and rotation angles separately using conventional ray tracing, i.e. Whitted ray tracing, using 64 samples per pixel to accurately simulate the blurring in curved regions due to limited camera resolution. Example input images are shown in Figure 5.7. We use a camera resolution of $2{,}048 \times 2{,}048$ pixels and simulate $\kappa_s = 2$ screens which results in $\kappa_s \cdot \kappa_c \cdot \kappa_r = 2 \cdot 11 \cdot 24 = 528$ light maps. Figure 5.8a shows the ground truth Stanford bunny model [TSDSR] from different viewpoints and Figure 5.8b shows the corresponding views on the reconstructed model. Furthermore, the visualization



(a) Camera image.

(b) Reconstruction.

Figure 5.6: *Results on a photometric stereo dataset: In particular, the painted regions of the clay mask exhibit specularities which leads to a violation of the assumption regarding Lambertian reflectance behavior. Nevertheless, the reconstruction preserves the shape in these regions. The robustness of our reconstruction technique originates from the outlier-robust normal estimation achieved by the mean-shift algorithm and a subsequent robust refinement step.*

of the Hausdorff distance between the original mesh and the reconstructed model is shown in Figure 5.9. The reconstruction fits to the original model except for the bottom region. The reason for the deviation at the bottom is that almost no information has been captured there. This is due to an inaccurately placed plane for the screen which illuminates the object from the bottom (see Figure 5.7). As a result, a small part of the feet of the bunny is not visible in the synthesized images.

In order to evaluate the robustness of our approach with respect to interreflections, we also consider a synthetic mirroring block with pits of increasing depth, where the proportion of multiple-bounce observations gradually increases. The reconstruction results are shown in Figure 5.10. In the deepest pit, the reconstruction deviates more from the ground truth than for the remaining parts. The reconstruction in such deeper concavities might be improved to some degree by using smaller turntable rotations than the used increments of $15°$, using more illuminants at different positions and by considering multiple reflections of the light rays.

Finally, we evaluate our technique on two mirroring real-world objects, again using turntable rotations of $15°$. To obtain information about the accuracy of our reconstruction approach, we have measured a precisely manufactured sphere with

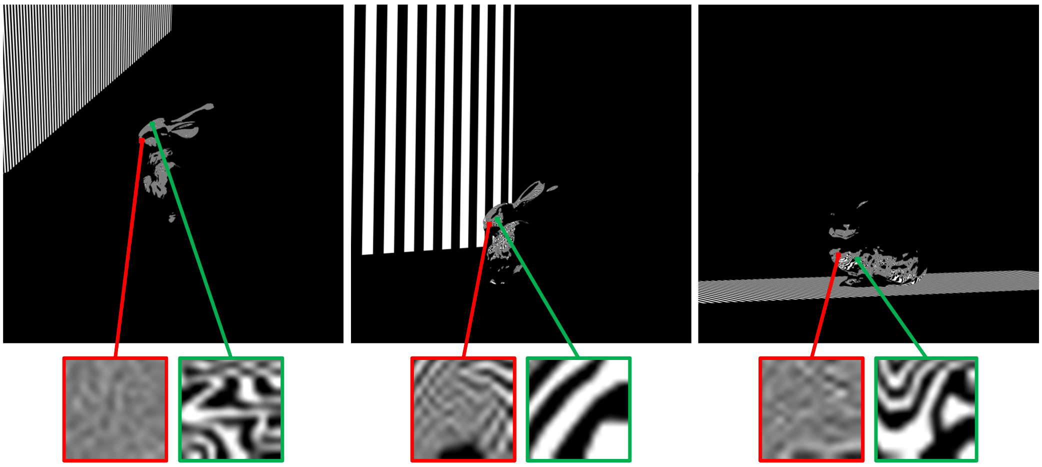


Figure 5.7: *Examples of the input images acquired by the cameras in the synthetic environment: The object is illuminated by one screen from the side (left, middle) and by one screen from the bottom (right). The images demonstrate the fact that the pattern appears blurred at some parts of the object surface due to local surface curvature and differing relative distances to the display pixels.*

a radius of 25mm (using $\kappa_s = 2$ screens) and compare the reconstructed model to an ideal sphere of fixed radius whose center is determined via a least squares fit. The error is measured via the Hausdorff distance and shown in Figure 5.11. The root mean square error of the reconstructed model is 20μm. Furthermore, we also compute the accuracy according to [SCD$^+$06]. This measure denotes the distance for which 90% of the vertices are within a certain distance d to the ideal model. For the sphere, we obtain a distance of approximately $d = 30\mu$m. Both accuracy measures are considerably lower than the edge length of a voxel (approx. 200μm) on the utilized maximum octree level nine and the approximately 150μm to which a single image pixel corresponds at the distance of the object. This sub-voxel accurate reconstruction results from the final refinement step applied after the continuous min-cut technique where local normal information is used to adjust the reconstructed surface in a narrow band with a width of one voxel.

In another test, we consider an object with a more complex geometry to test our technique in the presence of self-occlusions and concavities. For this purpose, we have acquired a mirroring bunny figurine. The acquisition process is similar to the one used for the previous experiment with the sphere. In more detail, eleven cameras, three screens, rotation angles of $15°$ and a maximum octree level nine are used. Figure 5.12 shows example images obtained during the acquisition process. In Figure 5.13, we illustrate the calculated consistency values and the obtained divergence values in a slice through the center of the volume. The reconstruction result shown in Figure 5.14 clearly indicates the possible reconstruction accuracy. Using additional octree levels could further improve the reconstruction but comes at the costs of a higher computational effort and higher memory requirements. More details about the individual experiments can be found in Table 5.1.

Table 5.1: *Details of the different experiments: The reconstruction of the Happy Buddha model [TSDSR] has been computed from input data consisting of per-view normal maps generated via an OpenGL normal shader. For the photometric stereo experiment on the clay mask, we used 198 point light sources evenly distributed in the upper hemisphere. The parameter κ_c denotes the number of cameras, κ_s denotes the number of screens, κ_r denotes the number of rotations, κ_v denotes the total number of the different views and κ_l denotes the number of the light maps utilized for the individual scenarios.*

Properties	Happy Buddha	Mask	Stanford Bunny	Block	Bunny figurine	Sphere
κ_c:	75	11	11	11	11	11
κ_s:	–	–	2	4	3	2
κ_r:	–	24	24	24	24	24
κ_v:	75	264	264	264	264	264
κ_l:	75	264	528	1,056	792	528

During our experiments, we usually start with an initial subdivision on level seven and each time perform three surface adaptions before going to the next higher octree level. On an Intel Xeon E5654 CPU with 2.4GHz, our level nine reconstruction for $\kappa_r = 24$, $\kappa_c = 11$ and $\kappa_s = 3$, as used for the real-world bunny figurine, which leads to 792 observed individual normal fields, requires approximately 12 hours, while the acquisition took approximately 2 hours.

The results shown in this section indicate the potential of normal-based surface reconstruction. In contrast to the previously presented multi-view normal field integration approaches in [CLL07, Dai09], our method is robust enough to deal with real-world data in the presence of noise and outliers. However, regions such as concavities with a certain orientation to the displays, under which no information can be observed, cannot be accurately reconstructed.

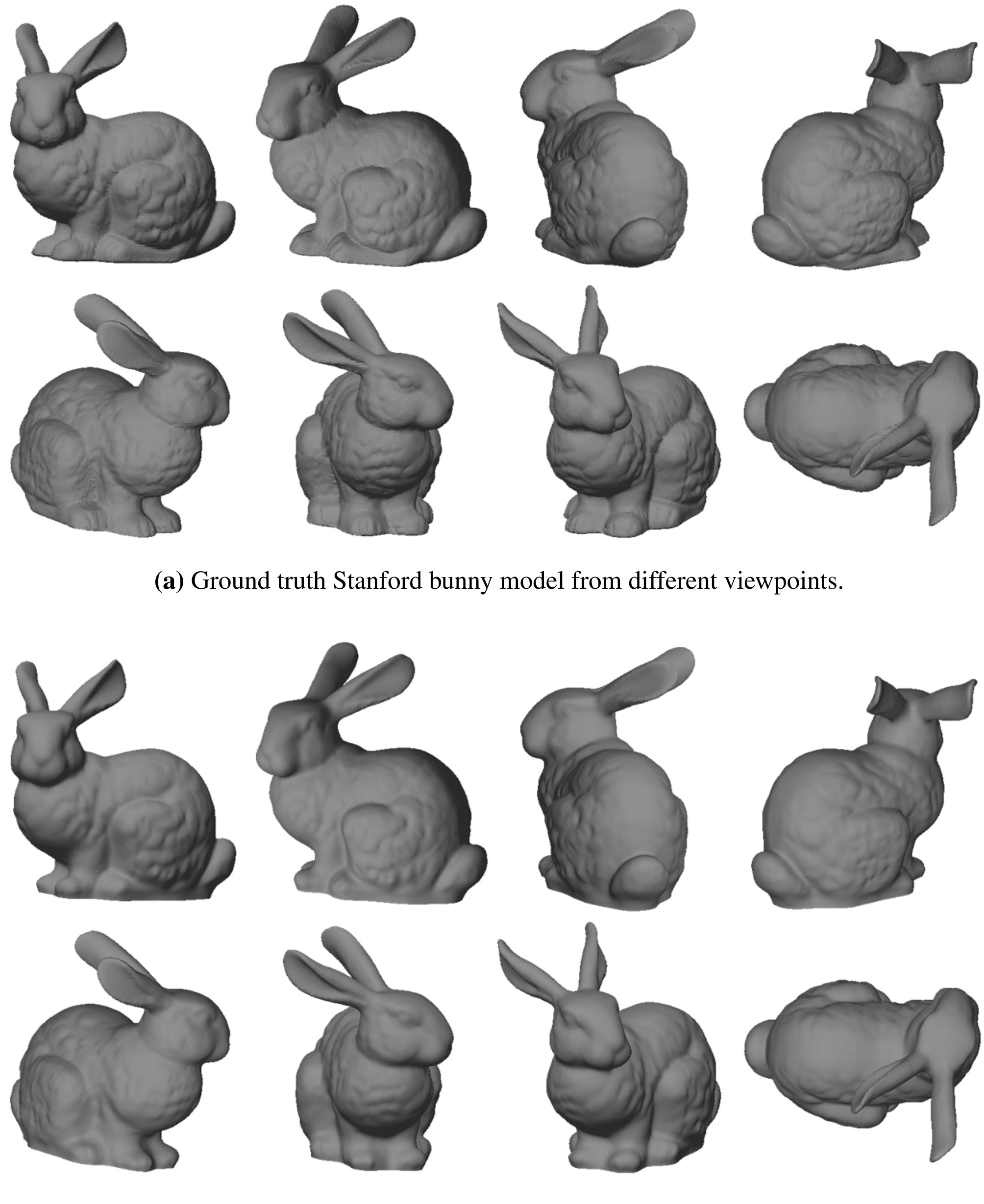
5.6 Conclusions

In this chapter, we have presented a novel, robust multi-view normal field integration technique for the reconstruction of the full 3D shape of mirroring objects. Based on coded illumination, our technique derives several normal hypotheses for each point of the considered volume. From these hypotheses, both the most likely local surface normal and a local surface consistency estimate are computed.

In our experiments, we have demonstrated that our method yields accurate 3D reconstructions of highly-specular objects even in the presence of occlusions. Therefore, our technique complements the investigations carried out in the previous chapters for objects with a surface reflectance behavior ranging from diffuse reflectance to opaque reflectance to specular reflectance with a sufficient diffuse component.

Currently, limitations of our technique can be found when considering deep concavities or other parts of the surface, where no information has been observed. Resolving these problems is challenging as it would require considering multiple scattering, i.e. multiple reflections of the respective light rays on their path from their origin on the light emitter to the camera.

Since the underlying optimization technique is independent from the source of the estimated normals, it would be possible to extend our method to objects which are only partially mirroring and also exhibit other surface reflectance behavior. This will be discussed in Part IV.



(a) Ground truth Stanford bunny model from different viewpoints.

(b) Reconstruction of the Stanford bunny model from different viewpoints.

Figure 5.8: *Ground truth Stanford bunny model and its reconstruction from different viewpoints.*

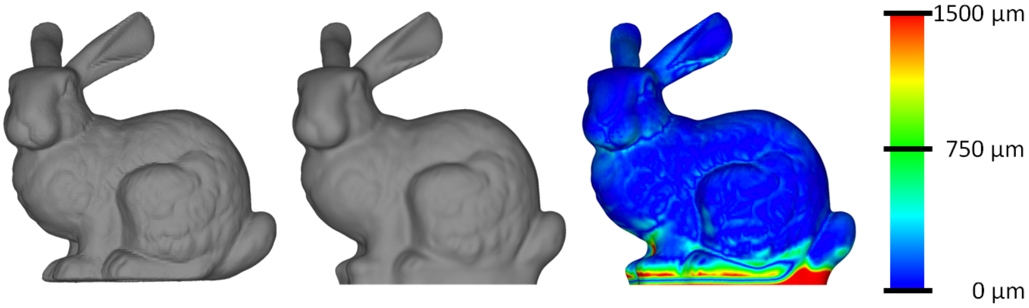


Figure 5.9: *Stanford bunny model, reconstructed model and visualization of the reconstruction error.*

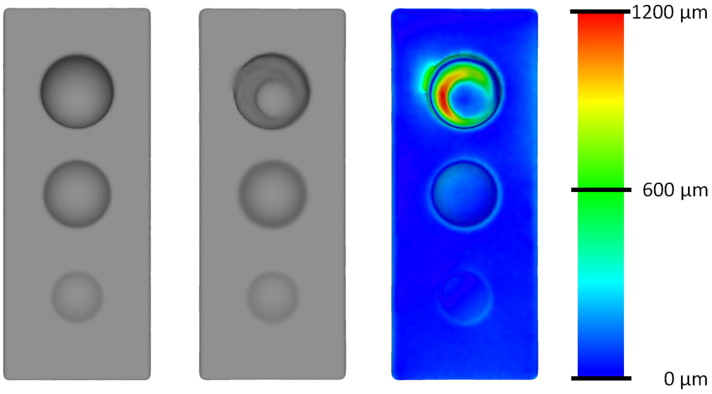


Figure 5.10: *Block model with pits of increasing depth, reconstructed model and visualization of the reconstruction error (level eight reconstruction, i.e. the voxel edge length is approximately $500\mu m$).*

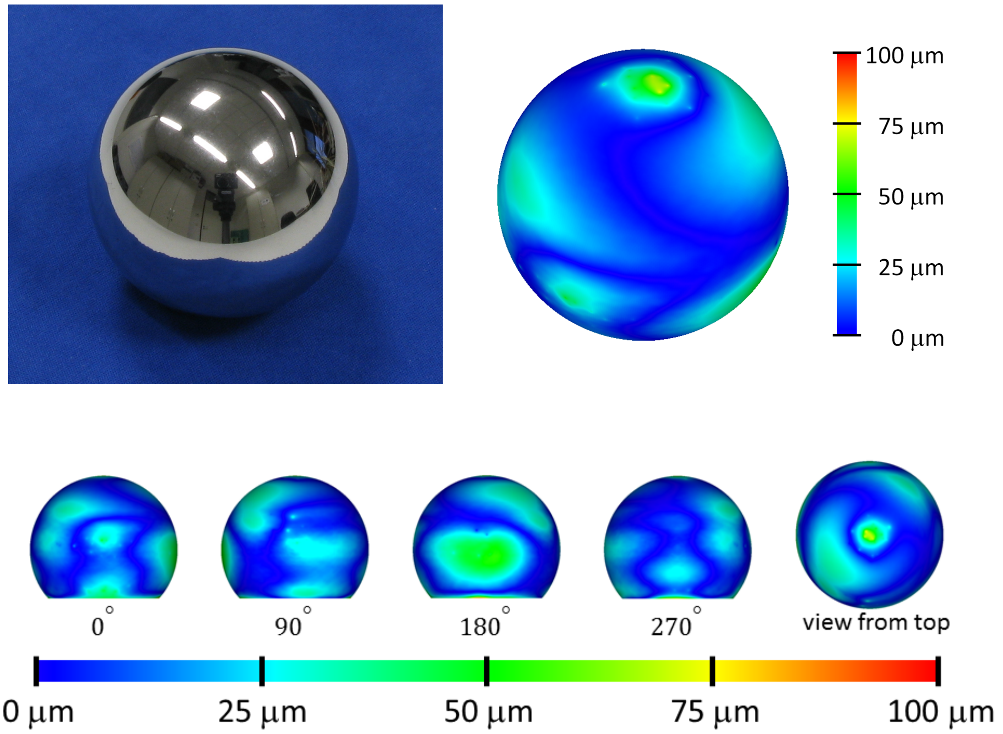


Figure 5.11: *Visualization of the Hausdorff distance for the reconstruction of an accurately manufactured mirroring sphere. The illustration shows the Hausdorff distance of the reconstructed mesh with respect to the ground truth for several views around the sphere as well as from the top view. As the sphere was placed on a non-mirroring sampleholder there is no data recorded for the bottom part of the sphere. The maximum deviation is reached at the top where there are almost no observations (level nine reconstruction, i.e. the voxel edge length is approximately* $200\mu m$*).*

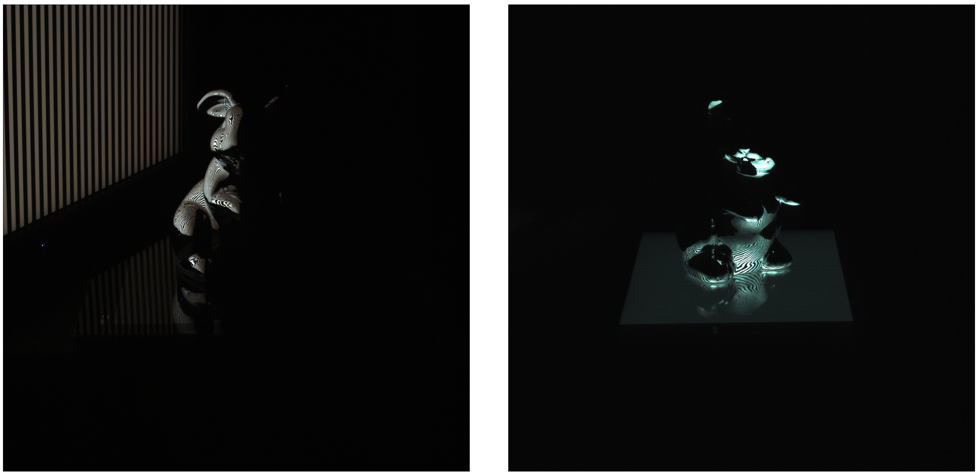


Figure 5.12: *Examples of the input images acquired by the cameras: The object is illuminated by one of the screens from the side (left) and by the tablet from the bottom (right).*

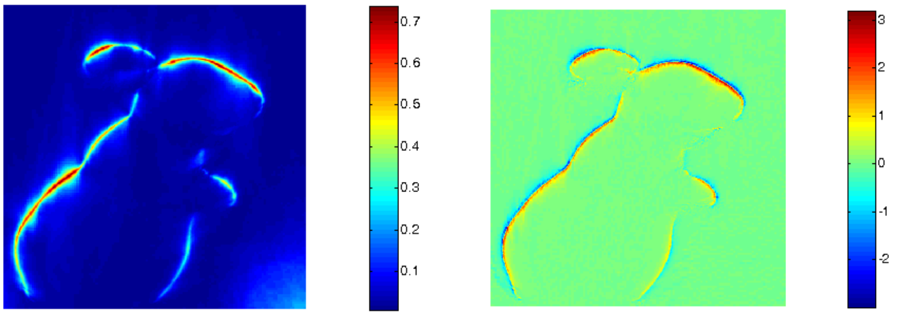


Figure 5.13: *Example slices for surface consistency and divergence for the mirroring bunny figurine: The slices are taken at an axis-aligned, vertical plane through the center of the volume of interest that is used to define the grid-based volumetric representation. The color values encode the consistency and divergence values respectively. At parts where information is observed during the acquisition, the surface is clearly localized. Smaller holes can be handled by the reconstruction approach.*

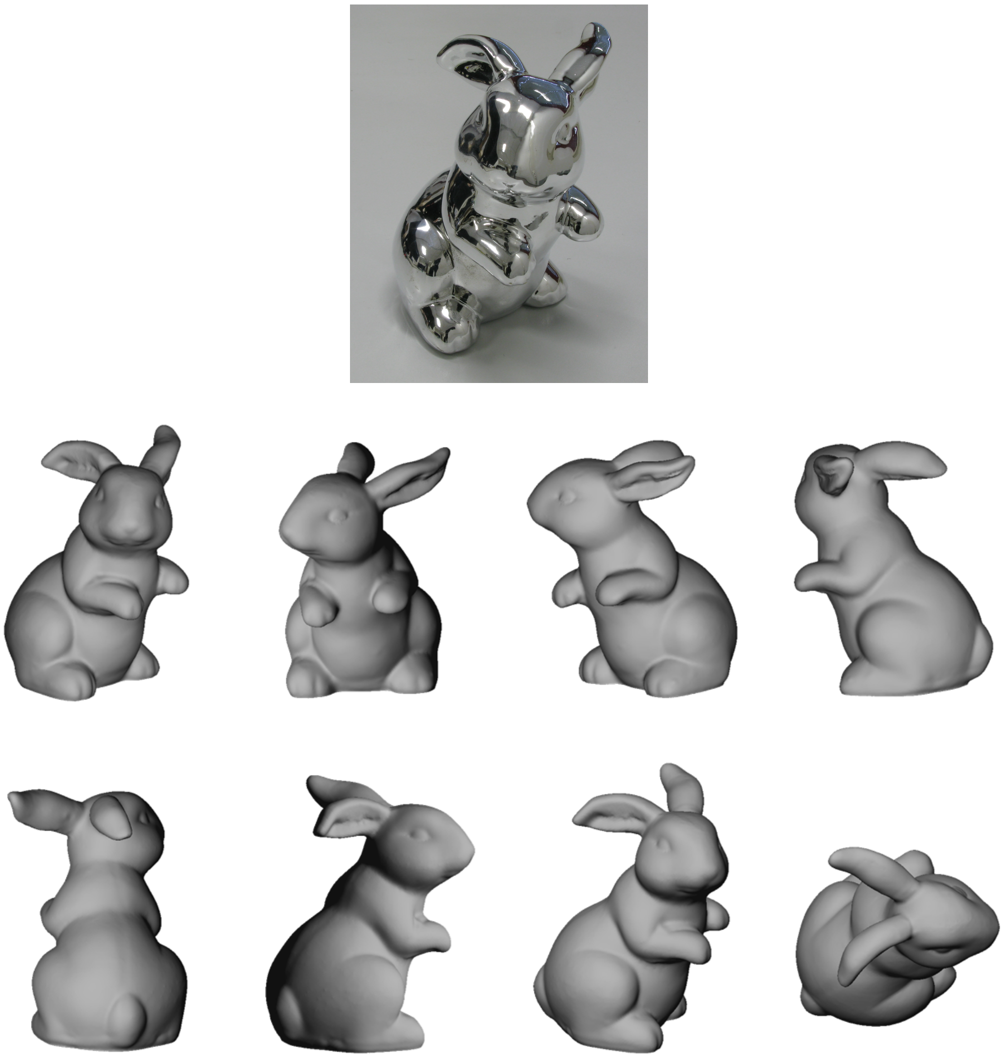


Figure 5.14: *Photo of a mirroring bunny figurine and respective reconstruction results.*

Part III

Image-Based Inference of Material Characteristics

CHAPTER 6

PRELIMINARIES

In order to analyze material appearance, it is essential to have a closer look at material properties such as color, texture, glossiness, translucency or transparency and how these can be visually derived from the image content. As mentioned in Chapter 2, the visual complexity of surface appearance is characterized by the complex interplay of surface material, surface geometry and illumination. For this reason, human perception can only observe material appearance depending on all of the involved modalities of material properties, surface geometry and illumination conditions determined by the environment. Similarly, standard acquisition devices are only capable of capturing the coupling of the respective modalities, which consequently influences the results of image analysis such as extracted feature descriptors. Directly separating these modalities would require a-priori information regarding a subset of these modalities and, hence, turn out to be a chicken-and-egg problem.

In the scope of this chapter, we will first provide a brief overview on material attributes (see Section 6.1) where we consider attributes in terms of linguistics and also review some of the key studies obtained in the field of psychophysics. Finally, we provide a short overview on descriptors that might be used to capture the aforementioned material characteristics from image data in Section 6.2 and an overview regarding techniques for material recognition that have been presented in literature (see Section 6.3).

6.1 Characteristic Material Attributes

Despite the complexity of visual material appearance, the human visual system is highly reliable in perceiving the characteristics of individual materials. As a result, we are able to rapidly recognize almost all of the materials around us on different surface geometries and under varying illumination conditions [SRA09].

While we can easily distinguish the individual surface materials, we usually can also directly group them into semantic categories such as textiles, leather, plastics, stone or wood. But what makes material appearance so characteristic to the human visual system and what are the traits that could be used in an automatic material recognition system? In the following sections, we first provide a short review regarding the characterization of material attributes in terms of linguistics, which represents the key ingredient for the communication of material attributes between humans (see Section 6.1.1). This is followed by a review of results obtained in studies on material perception (see Section 6.1.2).

6.1.1 A Linguistical View on Characteristic Material Attributes

When approaching the characterization of visual material appearance and the classification of materials, a first evidence might be obtained from linguistics. Linguistics represents a codification of human existence onto a level that is commonly comprehensible without the need for detailed knowledge regarding the underlying physical processes which, in some cases, even might be unfathomable. In fact, we are offered a rich vocabulary for putting characteristics of visual material appearance into words. Materials can be flat, rough, soft, hard, single-colored, multi-colored, matte, specular to even mirroring, translucent, transparent, homogeneous, inhomogeneous, etc.. Some investigations such as [BK69], [BRL97] and [CMK$^+$14] focus on describing attributes by words. While the studies in [BK69] focus on describing colors, [BRL97] and [CMK$^+$14] consider such a concept to describe texture information. In order to derive a texture lexicon, the investigations in [BRL97] consider cognitive aspects of perceiving texture information and, finally, 98 words have been used to describe texture attributes: asymmetrical, banded, blemished, blotchy, braided, bubbly, bumpy, chequered, cobwebbed, coiled, complex, corkscrewed, corrugated, cracked, crinkled, crosshatched, crowsfeet, crystalline, cyclical, discontinuous, disordered, dotted, entwined, faceted, fibrous, fine, flecked, flowing, fractured, freckled, frilly, furrowed, gauzy, gouged, grid, grooved, harmonious, holey, honeycombed, indefinite, interlaced, intertwined, irregular, jumbled, knitted, lace-like, latticed, lined, marbled, matted, meshed, messy, mottled, net-like, non-uniform, perforated, periodic, pitted, pleated, polka-dotted, porous, potholed, random, regular, repetitive, rhythmic, ribbed, ridged, rumpled, scaly, scattered, scrambled, simple, smeared, smooth, smudged, spattered, speckled, spiraled, spotted, sprinkled, stained, stratified, striated, studded, swirly, twisted, uniform, veined, waffled, webbed, well-ordered, whirly, winding, wizened, woven, wrinkled, zigzagged. The investigations in [CMK$^+$14] focus on visual aspects of texture and therefore ignore several of these words that are e.g. related to the shape properties of the respective object or do not correspond to visual features and merged similar words into a single category. The resulting 47 attribute words are: banded, blotchy, braided, bubbly, bumpy, checkered, cobwebbed, cracked, crosshatched, crystalline, dotted, fibrous, flecked, freckled, frilly, gauzy, grid, grooved, honeycombed, interlaced, knitted, lace-like, lined, marbled, matted, meshed, paisley, perforated, pitted, pleated, polka-dotted, porous, potholed, scaly, smeared, spiraled, sprinkled, stained, stratified, striped, studded, swirly, veined, waffled, woven, wrinkled, zigzagged. In another investigation [SN13], the attributes fuzzy, shiny, smooth, soft, striped, metallic, organic, translucent, transparent, rough, liquid, woven and man-made are used. Indeed, the consideration of attribute-based representations has led to promising results in the context of material recognition [SN13, CMK+14]. In other domains such as face verification (see e.g. [KBBN09, KBBN11]), attributes have also been successfully applied, where 65 considered attributes such as black hair, blond hair, eye width, mouth closed, smiling, etc. are focused on describing face properties. However, attribute-based techniques for detection or recognition purposes require the availability of huge amounts of annotated data, which implies enormous manual work.

6.1.2 A Psychophysical View on Visual Material Perception

While the adjectives mentioned in Section 6.1.1 have been well-established in human communication of material appearance, the human visual system itself rather perceives statistics of optical phenomena of light exchange happening on the material surfaces. Based on such statistics of e.g. perceived color or texture, an object can be perceived as shiny or matte, flat or rough, homogeneous or inhomogeneous. Even in the case of not finding an appropriate verbal description for a certain material sample, the human visual system is still capable of relating given material samples to psychological concepts and establishing links to similar objects that we have seen before [FWG13]. Following an example given in [FWG13],

> *metal objects come in an enormous variety of shapes and sizes, from needles to manhole covers to helicopters, and yet we are able to group metal materials together and make inferences about new exemplars based on our experience with other members of the class,*

which supports the quote of Anaxagoras mentioned in Chapter 1. The capabilities of human visual perception and our knowledge allow us to usually even assign the correct material class when materials might rather have the characteristic look of materials belonging to other material classes and therefore appear different in comparison to typical instances belonging to the same class due to large intra-class variances. As mentioned in [FWG13],

> *the reflectance properties, shape, and mesoscale texture of a piece of*

> *limestone may be more similar to bread or sponge than to quartz or crystal, and yet we most probably group both limestone and quartz into the class of stones while sponge and bread are quite different categories.*

Focusing on the distinctiveness of different material classes in terms of their perceptual qualities, the distinctiveness of different perceptual qualities and the relationship between visual estimates of material qualities and knowledge about material categories, the studies in [FWG13] consider human ratings of materials for nine perceptual qualities: glossiness, transparency, colorfulness, roughness, hardness, coldness, fragility, naturalness and prettiness. The respective results show that the ratings of the subjects involved in the study agreed with the category definitions fabric, foliage, glass, leather, metal, paper, plastic, stone, water and wood. This indicates that material properties play an important role in material classification by the human brain.

Representing one of the most essential capabilities of the human visual system, the visual perception of materials has been extensively discussed in literature. Among the most studied characteristic material attributes are surface roughness, glossiness and translucency. In the following, we only provide a rather brief overview on respective studies and refer to the more detailed surveys given in e.g. [And11, TFCRS11, Zai11, Fle14].

Regarding surface roughness, studies such as e.g. [PDG$^+$08, PK02] have been dedicated to an analysis of the perception of characteristics of surface reliefs. Furthermore, the investigations in [HLM06] indicate that roughness perception is biased by illumination conditions. For a directional illumination perpendicular to the material surface, the roughness does not seem to be as strong as for directional illumination from flat angles with respect to the material surface.

When analyzing glossiness, several studies focused on validating the ability of humans to judge surface glossiness [NS98, FDA03]. The results in [FDA03] show that reliable judgments can still be delivered under different illumination conditions which follow the statistics of natural illumination. Further investigations consider how the perception of glossiness, or specular reflection respectively, is affected by binocular disparity and motion [BB90, HCP91, WFM08, DFY$^+$11]. In addition, the properties of specular highlights also influence the perception of glossiness [BP81, BSBM05, FTA04, KMA11, MKA12, TNM04].

For translucent materials, the light enters the respective object, travels inside and leaves the object at another surface location. Such effects appear e.g. for wax, plastics and several types of minerals. Obviously, the perception of translucency by the human visual system depends on the absorption characteristics of the respective materials and the thickness of the objects as well as on the homogeneity or inhomogeneity regarding their refractive index. According to [Fle14], it still remains unclear how humans can differentiate shading gradients resulting from opaque reflectance from gradients resulting from subsurface scattering effects. Furthermore, the relation between specular highlights and the non-specular reflections has been identified to provide important information regarding the translucency of objects in the investigations in [Mot10]. In particular, translucency seems to influence specular highlights less than the non-specular regions. Analyzing the capability of the human visual system to directly match refractive indices of solid materials has been investigated in [FJT11]. While the human visual system is capable of coping with this task in the respective experiments, the perception of translucency has been significantly influenced by the thickness of the materials and their distance to the background.

Regarding the perception of materials, the argumentation in [Fle14] claims that different levels of fidelity might be required for different tasks as some tasks only involve a categorical judgment about some material attributes such as hard or soft whereas other tasks as e.g. encountered in design processes for marketing require considering the full range of possible variations of the respective attribute. According to [Fle14], the visual processing might therefore be grouped into material categorization or material estimation. While material categorization is focused on associating material samples with specific class labels by exploring information of other class members, the key objective of material estimation can be identified in deriving specific material characteristics such as the degree of specularity or the degree of elasticity while taking into account subtle discriminations in the space of variations of a certain material attribute.

In addition, Fleming [Fle14] suggests that the human visual system does not rely on estimating physical material parameters but instead uses a statistical appearance model

> *which captures the natural degree of variations between samples in terms of easily measured appearance properties.*

This means that statistics of low-level and mid-level image features might guide our perception of materials and, hence, also influence recognition processes. In the scope of this thesis, we follow the theoretical concept described in [Fle14] for our implementations of automatic material recognition systems. In particular, the material representations used for material recognition as described in Part III are based on statistics of low-level and mid-level image features. Based on the characteristics of these statistics, a characteristic material "footprint" is generated for the respective material. Such statistics might be based on observations of highlights, interreflections, subsurface scattering effects, etc.. While these statistical descriptions might be fundamental for the derivation of the above-mentioned verbal attribute dictionaries used in human communication, the internal processes of the human brain regarding material perception and material recognition probably are rather directly based on the statistical descriptions [Fle14]. Consequently, key problems that have to be dealt with concerning material perception and material recognition can be identified in the development of suitable descriptors that capture the individual visual appearance characteristics of materials as well as in the development of a suitable statistical representation based on these observed descriptors.

6.2 Material Attribute Descriptors

In order to transfer the aforementioned rather theoretical concepts towards their practical application in automatic systems, it is essential to find suitable descriptions to represent the visual material appearance in the image domain. In particular, the characteristic material traits such as shininess, roughness or homogeneity are manifested in characteristic local visual features with certain statistics of colors or textural patterns which represent the characteristics that determine the appearance of real or abstract elements or objects in an image and, hence, make these abstract elements or objects distinguishable [Wei13]. Such local features might not only represent localized patterns that vary from their local neighborhood [TM08] such as corner-like, blob-like or edge-like regions, but also the location-invariant local characteristics of colors or textures. The key challenge can be found in designing respective descriptors to capture these characteristic visual features. Clearly, finding appropriate features represents a task-specific challenge and, indeed, this has been a topic of research for decades.

By definition, local features provide a representation of local characteristics of the elements or objects and their appearance in the respective scene. This includes distinctive physical entities such as characteristic points or physical edges. Characteristic points in the scene might occur e.g. at corners of windows, corners of objects, corners in the surface texture of objects, etc.. Physical edges might be present in e.g. the surface structure of many objects such as the transitions from black to white in a checkerboard, the grain of wooden objects or object boundaries which determine the silhouettes of the objects in photographs. Furthermore, blob-like regions that appear on many surface textures, surface structures such as e.g. bricks and the surface structure of leather or wallpaper might also represent characteristic features in a scene. Such features have the advantage that they represent certain distinctive aspects of the scene which can be localized accurately. If such corners, edges or blobs occur in several images and can be reliably matched across the involved images based on their local neighborhood description, they might e.g. serve to establish relations between the different images such as the estimation of the respective parameters of the involved cameras which includes their relative pose to each other. For instance, corners of a checkerboard or specific quick response (QR) codes represent local features commonly used for camera calibration [Zha00, MS13]. So far, different feature detectors have been introduced in literature in order to find corners (e.g. Moravec detector [Mor80], Förstner detector [FG87], Harris detector [HS88]), blob-like regions (e.g. Laplacian-of-Gaussians detector [Lin98, MS04], Difference-of-Gaussians detector [Low04]) or other characteristic structures (e.g. Maximally Stable Extremal Regions (MSER) detector [MCUP02]). For detailed discussions regarding feature detectors, we refer to the survey in e.g. [TM08].

Furthermore, it might be possible to consider such localized features for the detection of objects or materials in images, i.e. the local neighborhood around such local features might be evaluated. However, such a strategy only considers a small subset of the possibly occurring patch structures. Therefore, many investigations such as [LM99a, LM01, LSP06, VZ09, CG08, CG10, XHE$^+$10, LSAR10, SLRA13] and our techniques presented in Chapter 7 and Chapter 8 skip such sophisticated strategies for localizing distinctive features in the scene and, instead, rather use densely sampled patches to describe scenes, objects or materials.

In order to describe the relevant image contents, appropriate descriptors have to be calculated in local regions surrounding the locations extracted in the feature detection stage. Detailed surveys on local descriptors are given in [MS05, TM08, MT09, vdSGS10, Wei13]. Capturing the variations in both local color distributions and local gradient distributions allows to capture characteristics of material appearance. Obviously, for materials with rough surfaces, more local gradients can be observed than for smooth surfaces as there are more interreflections or self-shadowing effects. Also, the color distributions will exhibit a larger variance than for a homogeneous flat material surface, and possibly occurring local highlights are captured in both color and texture descriptors. It is therefore not surprising that local color descriptors and texture descriptors are amongst the most popular descriptors regarding individual characteristics of material appearance. In order to consider local color distributions, local color patches similar to the illustration in Figure 6.1a have widely been used in literature (e.g. [VZ09, LSAR10, SLRA13]) and are also used in the scope of our investigations in Chapter 7 and Chapter 8. Local texture descriptors such as the Scale-Invariant Feature Transform (SIFT) descriptors [Low04] (see Figure 6.1b), Histogram-of-Oriented Gradient (HOG) descriptors [DT05], Local Binary Patterns (LBPs) [OPM02], basic image features [CG08, CG10], sorted-random-projection descriptors [LFCK12] and filterbanks [LM99a, LM01, Sch01, FA91, Gab46] (see e.g. Figure 6.2 for an illustration of the Leung-Malik filters, Schmid filters and Root Filters) focus on considering the local gradient information of the image intensities. Such descriptors or respective variants have widely been used to represent materials. In our techniques in Chapter 7 and Chapter 8, we also use such standard texture descriptors. Some of these descriptors such as SIFT and HOG have also been used in object classification, scene classification and registration or reconstruction respectively.

Independent from the respective task, an alternative to the standard descriptors discussed so far can be identified in descriptor learning strategies as e.g. proposed in [FZ07, BBS10, BHW11, SVZ12, OBUvdS13, SN13, SVZ14, DJV$^+$14, JSD$^+$14, LF14, BUSB14]. Such approaches focus on learning features that are capable of capturing the characteristics of the individual categories. This requires having adequate datasets which capture the appearance variations of materials, objects, scenes, etc., and, hence, might easily consist of hundreds of thousands of images. Therefore, these approaches are valuable if such datasets are available but rather impractical when the data does not represent the characteristics that are expected to be present in the query images used after the descriptor learning.

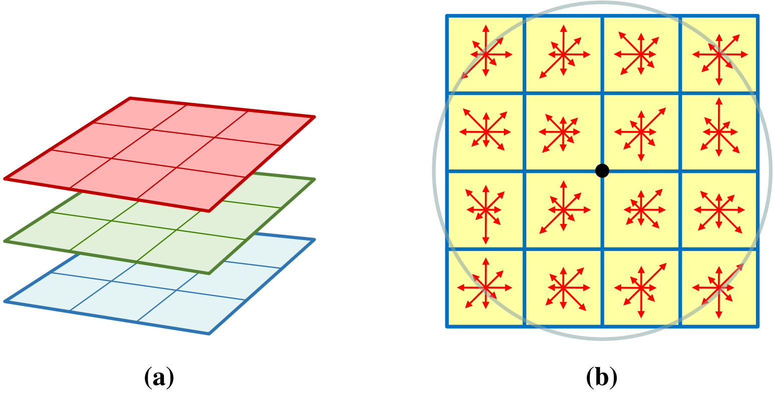
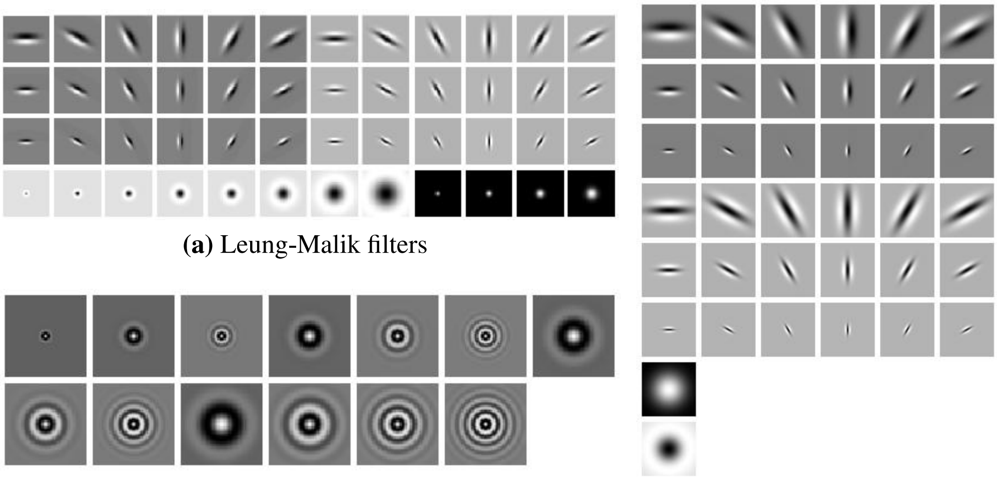


Figure 6.1: *Illustration of (a) the local color descriptor, where local color patches are considered and the intensity values in the 3 channels red, green and blue are concatenated to a* 27*-dimensional descriptor, and (b) the SIFT descriptor, where the local gradient information is used to generate histograms based on* 8 *orientations in a local* 4×4 *structure around the considered image point. Concatenating these values results in a* 128*-dimensional descriptor.*

While the aforementioned descriptor types do not directly correspond to semantic attributes of the image content as mentioned in Section 6.1, they can be used to derive more semantic material descriptors. In e.g. [CMK$^+$14], the Describable Textures Dataset (DTD) has been introduced as a new material attribute database for texture attribute recognition. Each of the 47 classes contains 120 images depicting the individual attributes in various forms and under varying viewing and



(**a**) Leung-Malik filters

(**b**) Schmid filters

(**c**) Root Filter Set

Figure 6.2: *Some of the commonly used filterbanks: (a) The Leung-Malik filterbank consists of edge, bar and spot filters at multiple scales and orientations. Its standard version includes two Gaussian derivative filters at six different orientations and three different scales, eight Laplacian-of-Gaussian filters and four Gaussian filters. (b) The Schmid filterbank includes* 13 *isotropic, "Gabor-like" filters and, therefore, provides a rotation invariant representation. (c) The MR filterbank corresponds to a common Root Filter Set of* 38 *filters. These filters include an edge and a bar filter at six orientations and three scales as well as a Gaussian filter and a Laplacian-of-Gaussian filter. However, only the maximum response across the orientations is used and, thus, only* 8 *filter responses are considered to establish the respective descriptor. Images taken from [Vis].*

illumination conditions. In particular, after computing per-image representations for each of the 47 involved texture attribute categories of the DTD, the responses of the respective attribute classifiers are trained. This results in a 47-dimensional vector which describes how well an image matches to each of the 47 attributes. Based on these trained models, images from other datasets are analyzed via the learned attribute classifiers resulting in a stronger representation. The approach presented in [SN13] is also based on considering attributes.

Similar to the aforementioned descriptor learning approaches, such techniques require capturing a huge multitude of different exemplars under a large variety of different viewing conditions, lighting conditions and surface geometries in order to appropriately represent the respective intra-class variation in appearance. Defining a single category appropriately might easily require several thousands of images. The required manual processes to capture exemplars as well as to segment and annotate materials in images severely limit the number of images per attribute category. This manual work is even more complex as many materials exhibit several of the above-mentioned attributes, which has to be taken into account in the annotations.

6.3 Material Recognition in Literature

Material recognition is a challenging problem due to the significant variations in material appearance under different configurations of viewpoint, illumination and surface geometry. In the following sections, we briefly discuss model-based approaches (see Section 6.3.1) and appearance-based approaches (see Section 6.3.2).

6.3.1 Model-Based Approaches

The inference of knowledge about the considered material surface can be approached by using certain models which describe the variations of material appearance under different view-light configurations. Histogram models have been used in e.g. [DN98] and [vGKD99] to represent the changes in appearance for materials under varying view-light conditions. In [OJR03], material recognition is approached based on a partly Lambertian and partly specular model. Furthermore, BRDF slices have been used for material classification in [WGSD09]. The studies in [CMPP02] and [DC05] analyzed the model-based dependency of texture features on illumination. However, such dependencies rely on certain surface characteristics, which also applies for the assumed reflectance models. While analytical models might be sufficient to represent the reflectance behavior of locally smooth surfaces with homogeneous reflectance behavior, they do not reflect characteristic material traits that determine the appearance of many materials with mesoscopic surface reflectance effects. Such effects take place at surface structures imaged to an area of approximately one pixel within an image. More complex material models such as bidirectional texture functions [DNvGK97] can deal with such mesoscopic effects but have their limitation with respect to extremely specular materials. In that case, their data-driven nature requires an ideally continuous angular sampling which would significantly increase the amount of data and is therefore rather impractical. Consequently, the selection of such model-based approaches is material-specific, i.e. the fitting procedures are guided by the appropriate model. In addition, the fitting involves the explicit consideration of a multitude of parameters such as the parameters of the reflectance model, the lighting, etc..

6.3.2 Appearance-Based Approaches

The key components of appearance-based material recognition systems are the extraction of discriminative descriptors that reflect characteristic material traits, an efficient and appropriate modeling of the material categories and an appropriate classifier. The prominent, often applied descriptors include color patches (e.g. [VZ03, VZ09, LSAR10, SLRA13, WGK14]), densely sampled SIFT descriptors (e.g. [LSAR10, SLRA13, WGK14]), Local Binary Patterns (LBPs) (e.g. [CHM05, LF12]), kernel descriptors [HBR11], filterbanks (e.g. [LM99a, LM01, VZ02, VZ04, CD04, CHM05, CHFE10]) and combinations of several of these descriptors (e.g. [BG06, LSAR10, HBR11, LF12, SLRA13, WGK14]) as using complementary descriptors for material recognition has been demonstrated to lead to superior results. In our material recognition techniques introduced in Chapter 7 and Chapter 8, we also use combinations of such color and texture descriptors. Furthermore, learning descriptors for material classification has been investigated in [SN13, LF14, CMK$^+$14, BUSB14]. After extracting such descriptors for the images contained in the training set and the test set, the descriptors from the training set are typically used to calculate a dictionary of representative descriptors denoted as textons. This allows assigning all of the extracted descriptors in an image to the respective visual words in the dictionary to get a texton-based image representation as introduced in e.g. [LM99a] and [LM01] and also followed in e.g. [VZ02, VZ04, VZ09, LSAR10, LF12, SLRA13, WGK14]. The resulting texton-based image representations can then be classified using nearest neighbor classifiers, Bayesian frameworks [VZ04, LSAR10], Markov random field (MRF) classifiers [VZ03], support vector machines (SVMs) [HCFE04, CHM05, LF12, LYG13, WGK14], random forest classifiers [Bre01], etc..

While most investigations focused on single-image-based material classification, some acquisition devices also offer the possibility to easily acquire several images under several view-light configurations, which might significantly facilitate material classification. Obtaining a highly reliable classification is of great importance if further steps of the acquisition procedure depend on the reflectance behavior of the material classified before. In [LM99a] and [LM01], histograms have been concatenated to form a single vector for each particular material, which imposes that materials are represented by a fixed ordering of the configurations within the combined vector where all the individual image representations have to be carefully registered. Comparing materials based on these vector-based representations hence requires that exactly the same view-light configurations are considered in each vector with the same fixed ordering. In [CD04], bidirectional feature histogram manifolds have been introduced as an alternative that does not require the same view-light configurations for both reference and query data. However, having a sparse set of view-light configurations represents a problem to this approach, as the reference manifolds become coarsely sampled. In addition, linear interpolation between neighboring view-light configurations will result in additional sources of inaccuracies which increase with an increasing distance of the neighboring view-light configurations. In our method described in Chapter 7, we aim at classifying material instances using only a few images and make use of results from the face recognition domain. For efficiency, we focus on training-free, linear approaches as presented in [CT10]. In particular, our material recognition approach yields significantly better recognition rates than previous methods while using smaller numbers of view-light configurations.

Other recent approaches include learning optimal illumination for material classification [JSJ10], material classification based on learning coded illumination to directly measure discriminative features such as projections of spectral BRDFs [GL12, LG14] or learning discriminative illumination patterns and texture filters to directly measure optimal projections of BTFs [LYG13].

So far, most of the literature focuses on material recognition in controlled lighting conditions. The few techniques that focus on material recognition under natural illumination include the bag-of-words frameworks based on state-of-the-art descriptors [LSAR10, LYG13] as well as descriptor learning approaches based on convolutional kernels [SN13] or multi-scale sparse coding [LF14], descriptors based on describable attributes in combination with Improved Fisher Vectors (IFVs) and Deep Convolutional network Activation Features (DeCAF) [CMK+14] or deep convolutional neural networks [BUSB14]. While these approaches are based on collections of annotated and segmented images given by the Flickr Material Database [SRA09], OpenSurfaces [BUSB13] or the Materials in Context Database (MINC) [BUSB14], such annotations and segmentations represent a lot of manual work and are typically obtained by costly crowdsourcing services such as Amazon Mechanical Turk (AMT) [BUSB13, CMK+14]. Instead, our approach described in Chapter 8 avoids the need for manual annotation and segmentation by using synthetic training data for the classification of materials under natural illumination.

CHAPTER 7

MATERIAL RECOGNITION UNDER CONTROLLED ILLUMINATION

In the scope of this chapter, we consider multi-view material recognition under controlled illumination conditions as present in e.g. lab environments. While most of the respective investigations have targeted material recognition based on single images, it seems evident that using richer query information including several images of a particular material under different view-light configurations might increase the robustness of the recognition. From the few methods that focus on material recognition based on several images, some rely on the presence of exactly the same registered view-light configurations in both training data and query data. As different acquisition devices typically have different constraints regarding the mounting of the involved components, the measurable view-light configurations differ for the devices. Consequently, relying on the availability of exactly the same registered view-light configurations in both training data and query data is impractical. The approach presented in this chapter overcomes these aforementioned problems by forming characteristic material spaces that consider the characteristics of the individual materials captured in the set of view-light configurations available per material. Material recognition can thus be formulated as a comparison of the individual material spaces based on set-to-set distances. We will demonstrate that our approach already allows a reliable material recognition based on only a few view-light configurations. Besides being part of this thesis, the method presented in this chapter has also successfully undergone peer review (see [WK15b]).

After a brief survey on material recognition from multiple views and a discussion of the contributions presented in this chapter (see Section 7.1), we provide the technical details of our material recognition approach (see Section 7.2) and the results achieved (see Section 7.3) before concluding the chapter (see Section 7.4).

7.1 Motivation

The automatic classification of materials represents one of the key enablers for the automation of industrial supply chains as many tasks have to be carried out depending on the material properties. To give an example, the task of grasping objects might have to be adapted according to the fragility of the respective objects, i.e. a fragile object has to be handled very carefully while less attention can be paid for non-fragile objects. Industrial environments typically allow the use of expensive, rather task-specific hardware equipment and the conditions of the environment can be controlled in order to maximize the capabilities of the resulting system with respect to the respective task.

While most of the studies that have been conducted in this research field focus on material recognition from just a single image, several of the available acquisition devices are equipped with multiple cameras, and, hence, offer the possibility to easily acquire several images showing the material under different view-light configurations. Obviously, information from several view-light configurations includes a richer representation of characteristics regarding the surface reflectance behavior and, hence, when used as query data, should result in clear benefits regarding a highly reliable material recognition in comparison to using only a single image. For instance, effects such as interreflections or self-shadowing might be less visible in a certain image. Using images taken under different view-light conditions captures these effects in a better way.

Material recognition based on multiple query images and multiple training images for the involved materials can be formulated as a set-based recognition task (see Figure 7.1). The per-image representations of the material appearance captured by certain material descriptors under the different viewing and lighting conditions form a kind of material space that is characteristic for the respective materials. Recognizing the closest material in the database to a certain query material can be achieved by comparing the individual material spaces. Evidently, information about material appearance under the different view-light configurations has to be represented in an adequate way for robust recognition systems. Seminal work, such as the investigations in [LM99a, LM01], is based on the availability of exactly the same registered view-light configurations in both training data and query data. The information regarding the appearance of a certain material under the different view-light configurations is represented by a single vector resulting from the concatenation of the per-image representations of material appearance. However, materials with representations obtained by using different view-light configurations cannot be compared which renders such a strategy impractical. Typically, the view-light configurations of different acquisition devices differ due to varying constraints regarding the mounting of the involved components. To overcome this limitation and allow for unregistered per-material query data consisting of images taken under different, unknown view-light conditions, our investigations focus on modeling material spaces using convex hulls and affine hulls as already successfully applied in the face recognition literature [CT10]. We demonstrate that material recognition can be achieved with a high reliability by looking at the characteristic material appearance under a few viewpoints. At first sight, this problem might seem to be not that interesting anymore due to the successful studies on databases such as the CUReT database [DvGNK96]. However, those databases offer only a small intra-class variance in the appearance of the involved material samples. Recent, more challenging databases with larger intra-class variances of the respective material samples such as the ALOT database [BG09] and our UBO2014 database used in the scope of this chapter and Chapter 8 have shown that there is a need to obtain further insights into recognizing materials using multiple view-light directions for the reference/query sets.

In summary, the key contributions of this chapter are:

- a novel, robust framework that allows the recognition of materials under controlled illumination conditions based on sets of images which are acquired under different view-light configurations, and
- a study for using set-based classifiers to find the closest material in the database from a set of view-light configurations which might not necessarily be contained in the database.

7.2 Methodology

The basic principle of our approach is illustrated in Figure 7.2. For a query material, we search its best representative within a material database which contains images of a multitude of material samples taken under different viewing and lighting conditions that are expected to be met during the acquisition with standard devices such as the ones discussed in [SSW$^+$14].

For a reliable material recognition, we need to consider the spatial variations of a material as well as its change in appearance induced by different viewing and lighting conditions. Therefore, our approach is based on first computing state-of-the-art descriptors to capture the characteristic material traits and the subsequent derivation of a vector-based representation for each of the given images under individual viewing and lighting conditions (see Subsection 7.2.1). The set of vectors resulting for an individual material sample is then used to obtain its material space. This allows us to perform the comparison of different material spaces via set-to-set distances (see Subsection 7.2.2).

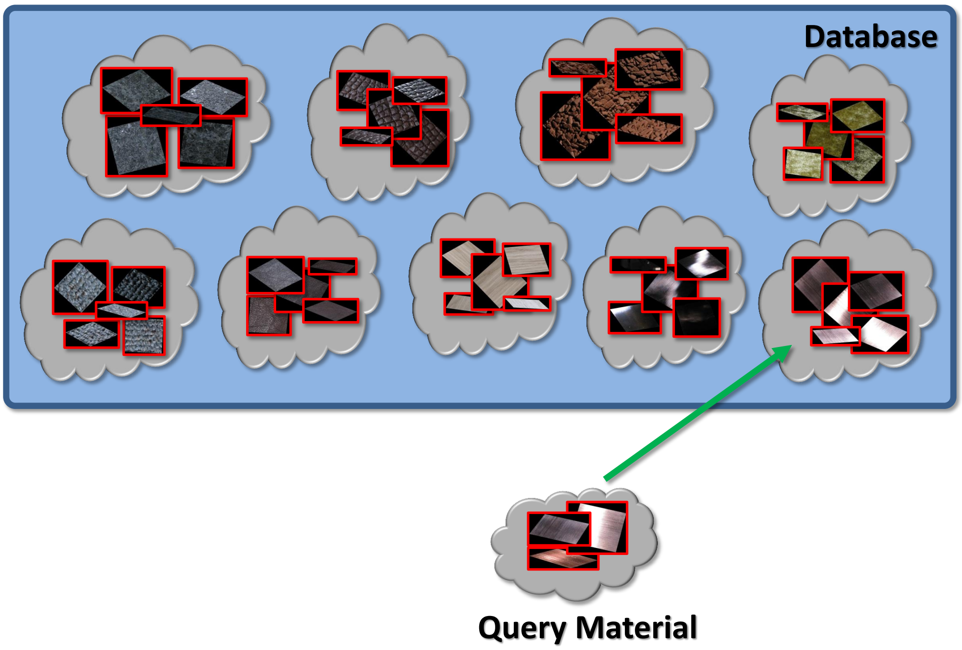


Figure 7.1: *Formulation of material recognition using multi-view information as a set-based recognition task: Characteristic material traits observed in the images of a particular material instance form a characteristic material space. The objective is to identify the most similar material instance within the database for an input query material by comparing the material spaces.*

7.2.1 Material Representation

In order to obtain a representative model per material, material-specific properties have to be included in the set-based representation. Characteristic material traits can be identified in a huge number of different aspects such as color, surface roughness, self-occlusions, interreflections, or specularities, and it has been shown to be beneficial to use several feature descriptors considering different types of attributes (e.g. [LSAR10]). We consider the following descriptor types which are densely sampled on a regular grid with a spacing of 5 pixels in our experiments:

- Color: We extract 3×3 color patches and concatenate the respective entries to a vector as in [LSAR10] which results in a 27-dimensional representation.
- SIFT: In order to consider the local spatial and directional distribution of image gradients, we extract dense, 128-dimensional SIFT descriptors as in e.g. [LSAR10]. SIFT descriptors provide robustness to variations in illumination and viewpoint.

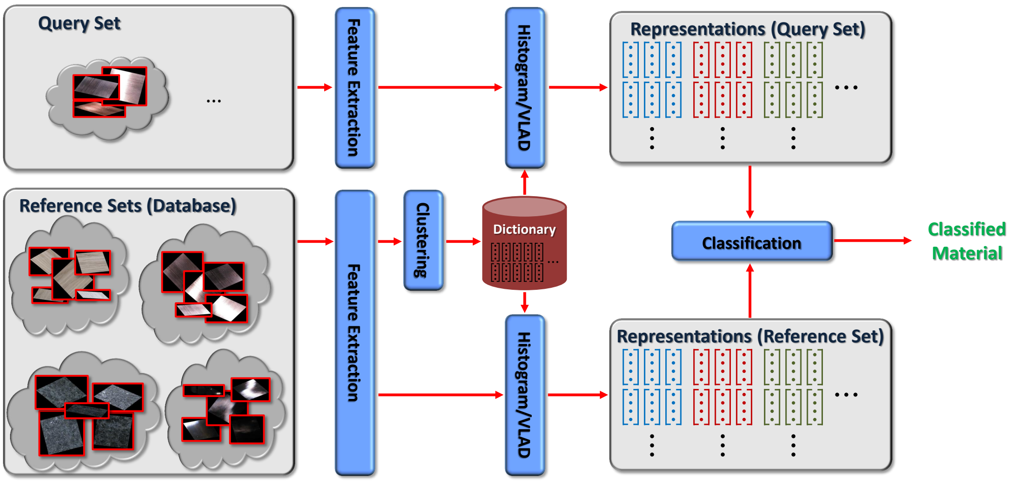


Figure 7.2: *Set-based material recognition scheme: After extracting descriptors, we compute a dictionary from the descriptors obtained for the reference data. This dictionary is used to quantize the representation of the content of a particular image into a vector representation. Finally, a set-to-set classification is carried out to find the closest material within the database.*

- HOG2x2: After computing histograms of oriented gradients, neighboring descriptors are concatenated to a 124-dimensional descriptor as in [XHE+10]. As the normalization differs from the scheme used in SIFT, it captures material characteristics in a different way.
- Leung-Malik filters: LM filters [LM99a, LM01] represent a filterbank which consists of edge, bar and spot filters at multiple scales and orientations. In our experiments, we use two Gaussian derivative filters at six orientations and three scales, eight Laplacian-of-Gaussian filters and four Gaussian filters. The filter responses obtained per pixel when applying these 48 filters onto an image form the corresponding 48-dimensional descriptors.

The extracted descriptors are then used to compute a vector-based representation for each of the image regions that show the respective material. In the scope of this chapter, we analyze the suitability of the popular bag-of-words representation and the more sophisticated VLAD representation [JDSP10]. Based on the descriptors extracted from the images contained in the database, we compute a dictionary of visual words for each descriptor type via k-means clustering [Ste57, Llo57, Llo82, Mac67] due to its efficiency. Please note that other clustering schemes might also be applied. In case of the bag-of-words model, we quantize each descriptor to its closest visual word in the dictionary and form histogram representations. In contrast, the VLAD representation is based on first assigning all local descriptors

$\mathbf{x}_i$ within an image region to their nearest neighbor $\mathbf{c}_j$ with $j = 1, \ldots, k$ in the corresponding dictionary with k visual words for each feature type. Then, the VLAD entries are computed by accumulating the differences of the local descriptors and their assigned visual words following

$$\mathbf{v}_j = \sum_{\{\mathbf{x}_i | \mathrm{NN}(\mathbf{x}_i) = \mathbf{c}_j\}} \mathbf{x}_i - \mathbf{c}_j. \tag{7.1}$$

These entries are concatenated to the final VLAD vector, which we normalize to unit length for each of the descriptor types. As a result, the VLAD representation incorporates information about the direction and the offset of the descriptors assigned per cluster. In Figure 7.3, we illustrate the calculation of both histogram representations and VLAD representations. These vector-based representations for the image regions of a particular material instance acquired under different view-light configurations form its corresponding material space. When combining several descriptor representations, we simply concatenate the normalized vectors corresponding to the involved descriptor types.

7.2.2 Set-Based Recognition

In contrast to e.g. [CD04], where a-priori knowledge about the considered viewing and lighting conditions is incorporated for setting up aligned training manifolds, our approach does not rely on the availability of such information. A randomly taken subset of images without knowledge about the imaging parameters should be enough to reliably recognize materials. We use the linear methods presented in [FY03, YFM98] and [CT10], where there is no need for parameter learning. Non-linear techniques (e.g. [CT10]) could be employed as well at the cost of learning hyper-parameters such as the kernel width.

Linear convex hull based classifier: Representing the material instances via vector representations for the respective images acquired under different view-light conditions allows us to make use of the convex hull classifier presented in [CT10]. Here, we assume that the vector representations under the available view-light configurations chosen to represent one of the individual material samples can be represented via convex hulls. The distance between convex hulls can be calculated by using quadratic programming and is abbreviated via CHISD (Convex Hull based Image Set Distance) as in [CT10].

Linear affine hull based classifier: Similar to the studies in [CT10], we consider affine hulls to represent the material spaces. Here, the material representations

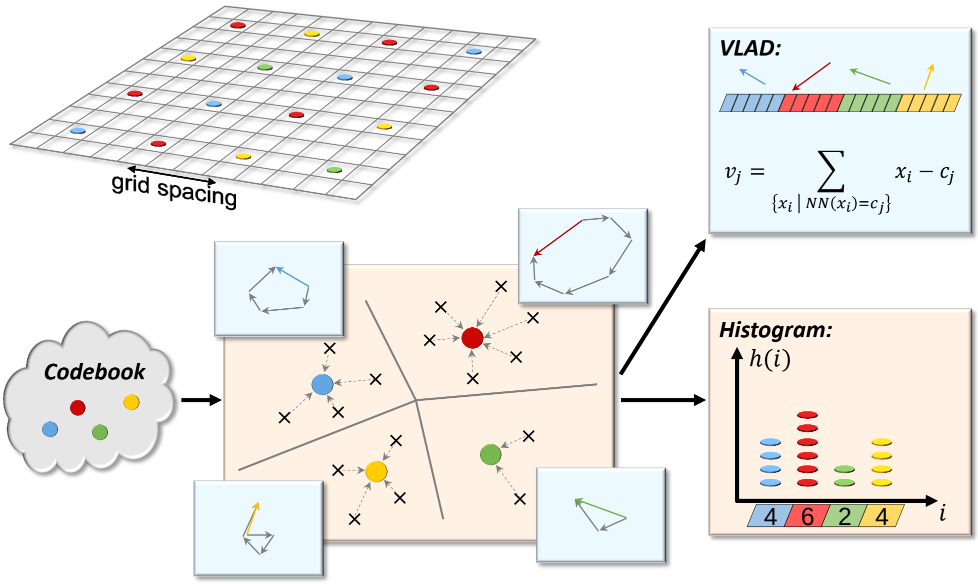


Figure 7.3: *Illustration of the computation of histogram representations and VLAD representations: The extracted descriptors are first clustered. For this purpose, we use k-means clustering [Ste57, Llo57, Llo82, Mac67]. However, other clustering schemes would also be possible. Subsequently, the histogram representation can simply be calculated from the assignment of the extracted descriptors to their closest cluster center. While we use hard assignments, i.e. a descriptor is completely assigned to only its closest neighbor, a soft quantization to several centers would also be possible. In contrast to the histogram representation, the VLAD representation [JDSP10] also incorporates both direction and magnitude of the offset of the individual descriptors to their closest centers.*

are obtained by affine combinations of the vector-based representations of the statistics of the material under the measured view-light configurations. We calculate the linear affine hull parameters by computing an orthonormal basis for the affine subspace spanned by vectors representing a particular material. The distance between two linear affine hulls abbreviated via AHISD (Affine Hull based Image Set Distance) can be computed by using the hyperplane which optimally separates the affine hulls.

Mutual subspace method (MSM): This type of method used in [YFM98, FY03] represents each class with a subspace formed by the respective vectors, and the similarity between subspaces is determined by comparing the angles between the subspaces.

7.3 Experimental Results

In order to compute the histogram and VLAD representations respectively, we used dictionaries with 150 visual words for color, 250 visual words for SIFT and 200 visual words for the LM filters similar to [LSAR10, SLRA13] throughout all of our experiments. For the HOG2x2 descriptors, we use 250 visual words.

In the scope of our experiments, we aim at analyzing the capabilities of the different set-based recognition techniques. We therefore perform experiments on different datasets for varying numbers of view-light configurations in the reference and query sets. We always take disjoint sets of view-light configurations for the reference/query sets of the material samples, i.e. images acquired under different view-light configurations are used for the reference set and the query set per material. In the following sections, we discuss the performance of our material recognition framework on several datasets.

7.3.1 CUReT Database

In order to obtain an intuition of the recognition performance, we use the well-established LM filters and denseSIFT to recognize the 61 material samples provided in the CUReT database (see Figure 7.4). Using 5 randomly chosen view-light configurations to represent both reference and query materials, we already obtain high accuracies of around 95.5% for both LM filters and denseSIFT when using AHISD and CHISD with VLAD representations. MSM methods perform worse by about 5%. The benefit of the high-dimensional VLAD representation becomes obvious in the fact that histograms perform significantly worse by $4\% - 11\%$. Using more view-light configurations to span the space for the different material samples, we observe that the accuracy obtained when using the individual descriptors closely approaches the 100% already for about 10 view-light configurations in reference and query sets. In general, there is a tendency that the high-dimensional VLAD description gives better accuracies than using histograms. We also combined the descriptors which further increases the performance.

In [CD04], a selection of 20 material instances of the CUReT database has been analyzed. Using 56 images per material instance for their reference manifolds, a performance of about 98% has been reached for the classification of individual textures and a bit more than 70% when using 10 view-light configurations per material. For a fair comparison, we only use LM filters as descriptors. When representing the reference sets with 10 randomly drawn view-light configurations and having a single configuration for the query material, we obtain performances of around 95% for the combination of CHISD and VLAD representations. This is only slightly worse than the 98% reported in [CD04] for reference manifolds based on 56 images per material instance. In a direct comparison with using 10 configurations for the reference sets, this combination of CHISD and VLAD leads to an improvement of about 20%. When using more configurations in the query sets, we already reach more than 99% starting from three view-light configurations per query set. A similar performance can be achieved with densely extracted SIFT descriptors.

The high performances reached on this database indicate that the individual material samples appear rather distinctive and that the database is not highly challenging. Additionally, as a consequence of the high performances, a real analysis of the different set-based methods with respect to each other is hardly possible and the need for set-based recognition is not yet clearly visible. For this reason, more insights can be obtained by using more challenging datasets with higher intra-class variances in material appearance under different view-light configurations.

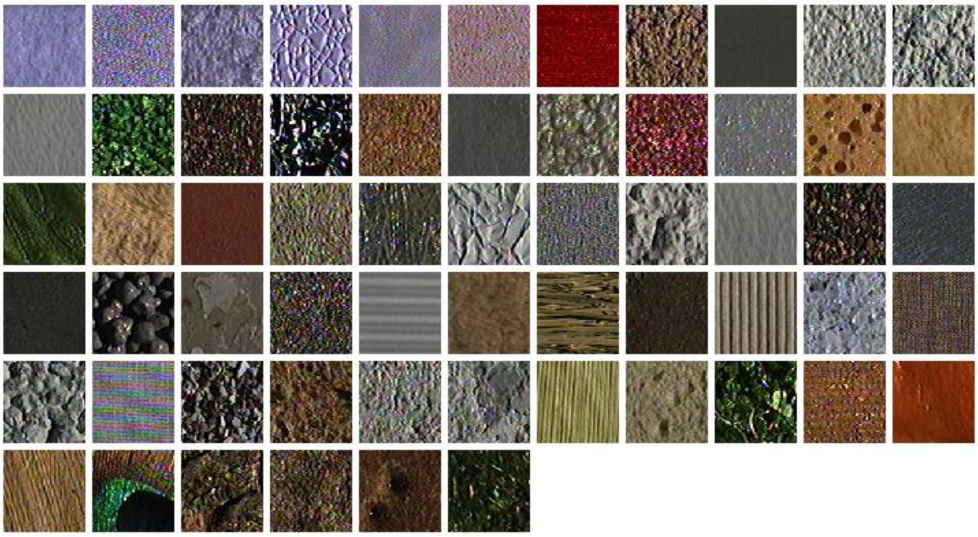


Figure 7.4: *Material samples in the CUReT database [DNvGK97] (taken from [Vis]).*

7.3.2 ALOT Database

The Amsterdam Library of Textures (ALOT) [BG09] (see Figure 7.5) offers significantly more and also a wider range of different material types, which have additionally been observed under different illumination colors. In our experiments, we consider color patches, denseSIFT, HOG2x2, LM filters and their combination.

Taking 5 view-light configurations for each of the reference and the query sets results in an accuracy of about 60% for color, 89% for denseSIFT, 83% for LM filters and 83% for HOG2x2 when using VLADs with CHISD or AHISD and 4% less when using VLADs with MSM methods. Furthermore, using histograms instead of VLADs generally leads to lower performances. The combination of the descriptors, however, leads to about 94% for AHISD and CHISD with VLAD and little lower accuracies for the MSM methods. Taking 10 configurations per reference/query set, the accuracies of the individual descriptors increase, and for the combination of descriptors we reach slightly above 99% using all the methods. This indicates a trend that the reliability of material recognition increases with increasing numbers of view-light configurations for reference and query sets.

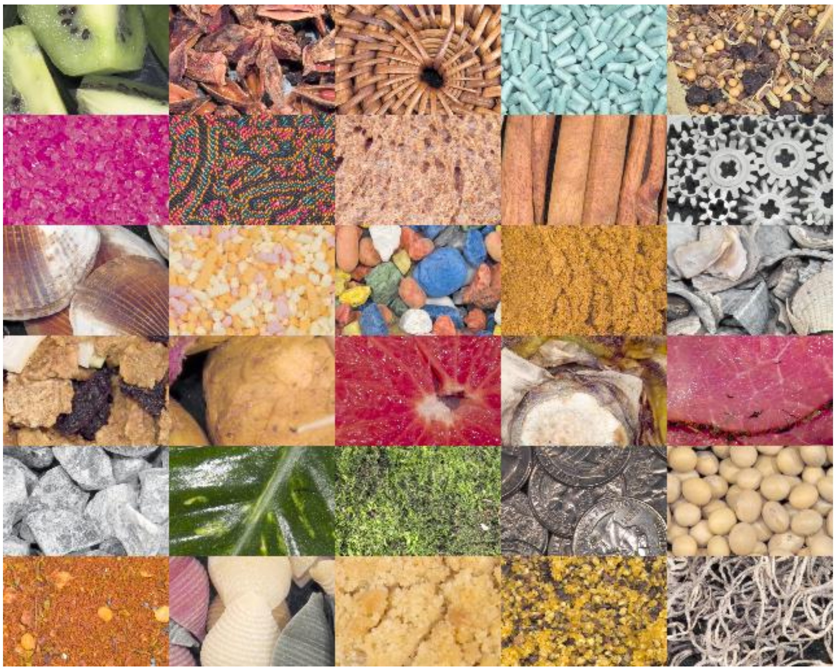


Figure 7.5: *Material samples in the ALOT database [BG09]. For illustration purposes, only a subset of the* 250 *material samples of the ALOT database is shown.*

7.3.3 Measurement Data of the UBO2014 Database

While the ALOT database [BG09] gives a more visible impression of the power of set-based recognition, the samples in this database still do not seem to show very extreme intra-class variations under different view-light configurations in comparison to the inter-class variances. In contrast, the material samples of the UBO2014 database that will be described in detail in the scope of Chapter 8 are used to model the variance in different semantic categories. We use the measurements of the 84 material samples as used in Chapter 8 and further 76 material samples in the database extension (see Figure 7.6). For each of these material samples, photos have been taken under 151 different viewing directions and 151 lighting directions, which leads to $22{,}801$ images per material sample. For some of the categories, several of the samples only exhibit rather subtle differences (e.g. tiles or metals). This makes the dataset challenging. Instead of grouping these samples into semantic categories as in Chapter 8, we consider the measurements per material sample individually and focus on recognizing the material samples. As illustrated in Figure 7.7 and Figure 7.8, the accuracy again increases if the number of configurations considered in the reference/query sets is increased. As before, we observe the trend of VLADs to be more discriminative than histograms. Furthermore, the descriptors have been evaluated separately, where denseSIFTs tend to perform best. The difference with respect to the performance of other descriptors is more visible for the histogram representations. AHISD and CHISD almost consistently outperform the MSM methods. Using the combination of different descriptors results in improvements over the accuracies obtained for the individual descriptors. These improvements are larger, if only a few configurations are available for the reference/query sets.

In Figure 7.9, we illustrate the dependency of the obtained accuracy on the number of view-light configurations in the query sets. We only depict this information for CHISD, which outperformed the other classifiers in the previous experiments. However, we additionally show the performances of using individual descriptors and some combinations of the descriptors. In general, the obtained accuracies increase when taking more images in the query sets, and using the VLAD representation leads to accuracies superior to the ones obtained when using histograms. Furthermore, we analyze the accuracies obtained by using different combinations of the descriptor types. The difference in the obtained accuracies indicates that the descriptor types carry different amounts of complementary information. In particular, the combination of color and denseSIFT clearly outperforms the remaining combinations of two descriptor types and even slightly outperforms the combination of all four descriptors types. Additionally, it becomes apparent that considering multiple view-light configurations leads to significant performance gains of almost 20% when using 10 configurations for the query sets in comparison to using a single configuration for the query sets when considering the combinations of descriptors. For more view-light configurations in the query set, we observe rather marginal improvements in the accuracies which indicates that the appearance variations that span the material spaces of the individual material samples are adequately captured. When analyzing the few misclassified material samples (e.g. two of the tiles as shown in Figure 7.10 and two of the metals have not been properly distinguished) and the respective estimated material labels, we observed that the estimated material and the ground truth material indeed look rather similar and it is even hard for the human eye to distinguish them.

7.4 Conclusions

In this chapter, we have presented a study on using set-based recognition schemes in combination with standard descriptors and encodings for material recognition. Our study demonstrates the benefit of making use of several images of a material sample for different view-light conditions regarding material recognition. There are only little performance gains for databases with smaller intra-class variance before reaching the saturation close to 100%, which might have led to less interest in investigations on material recognition based on several view-light configurations in recent years. However, when considering more challenging databases with larger intra-class variances under different view-light configurations, it is significantly more difficult to provide a reliable material recognition. We have shown that such a material recognition can be achieved with a high reliability by looking at the characteristic material appearance under a few view-light configurations which emphasizes the significant benefit of set-based material recognition in the presence of larger variations in appearance of the individual samples. Such a highly reliable material recognition technique represents one of the fundamental prerequisites of an efficient acquisition of geometry and reflectance as will be shown in detail in Chapter 9.

The approach presented in this chapter focuses on reliable material recognition under controlled illumination conditions. In addition, our method allows considering material information from several view-light configurations in the query data. These conditions might be given in typical industrial applications such as material inspection and are naturally given in standard material acquisition devices as e.g. mentioned in [SSW+14] in a lab environment.

Furthermore, our approach might also be interesting for multi-view material recognition in more complex, natural illumination. However, when considering such natural illumination scenarios, there are typically two fundamental changes. Firstly, the huge variation of illumination conditions needs to be taken into account when defining the individual material spaces. But the even more crucial difference is that the typical application scenario is based on completely different priors. Whereas a systematic acquisition of material information with special and expensive hardware equipment involving several view-light configurations is easily possible in controlled lab environments and makes sense regarding e.g. material inspection in industrial environments, the same task becomes highly challenging when considering the huge variations of different illumination conditions encountered in real-world scenarios which severely influence material appearance. Furthermore, this scenario is less determined by industrial hardware devices but more by lightweight, low-budget consumer acquisition devices such as simple cameras, cameras of a mobile phone or cameras of tablets. Therefore, in such a scenario, material recognition from single images seems to be the more natural scenario. Unfortunately, research is still in an infant stage regarding material recognition in such scenarios. We therefore leave the application and maybe required adaption of our set-based material recognition approach to uncontrolled illumination conditions for the future, and focus on the probably more typical and even more challenging single-image based material recognition task in the scope of the next chapter.

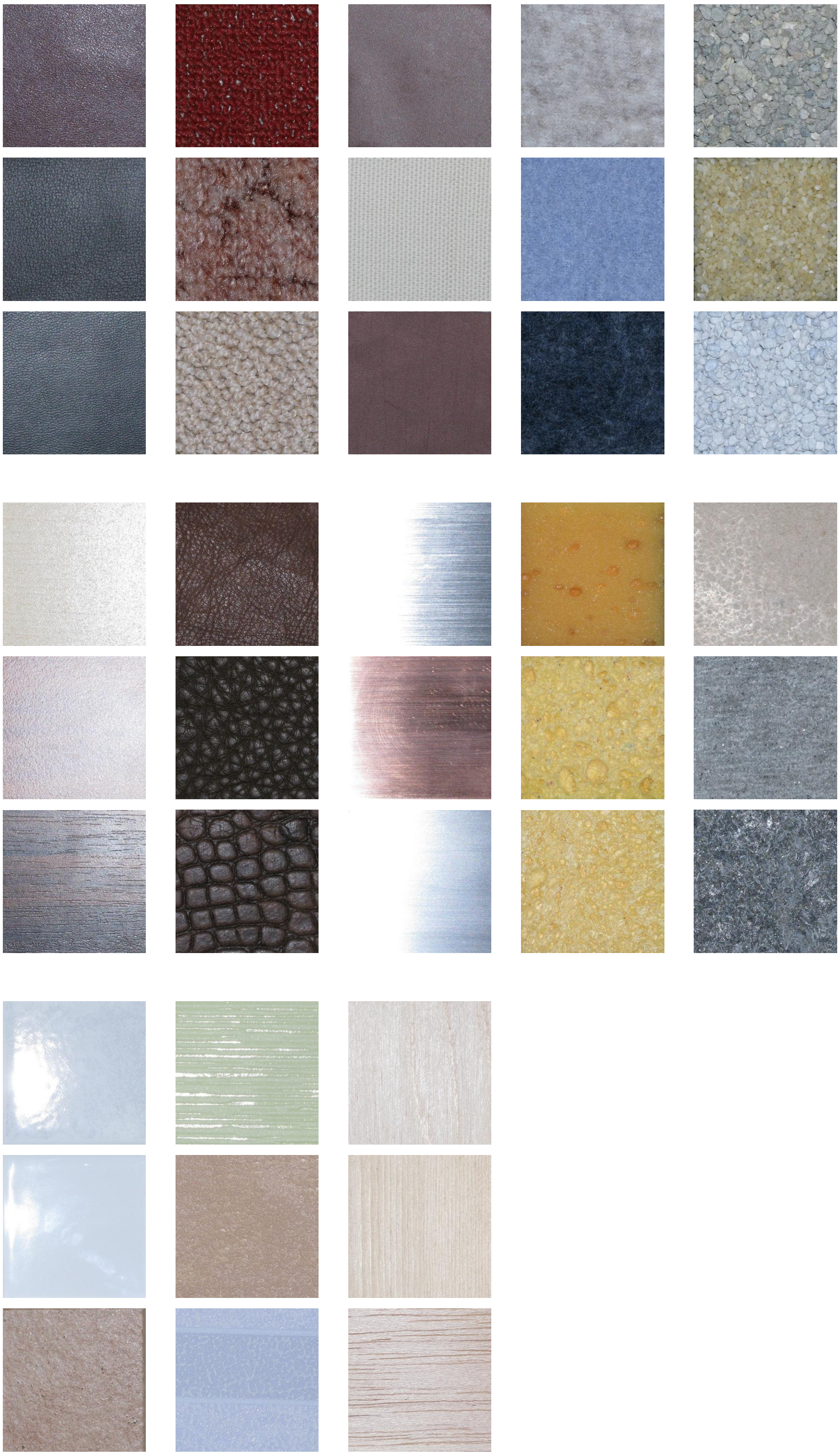


Figure 7.6: *Some of the materials measured for the UBO2014 database and its extension.*

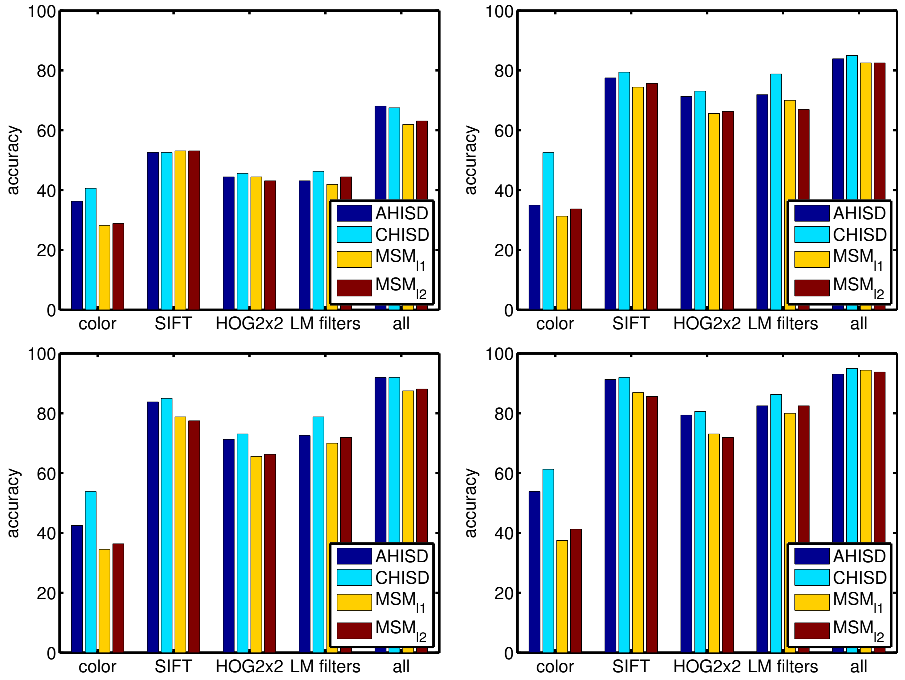


Figure 7.7: *Accuracies obtained for the data measured for the UBO2014 database and its extension when using histograms, disjoint reference and query sets of* 5 *(upper left),* 10 *(upper right),* 15 *(lower left) and* 20 *(lower right) randomly drawn images taken under different view-light configurations.*

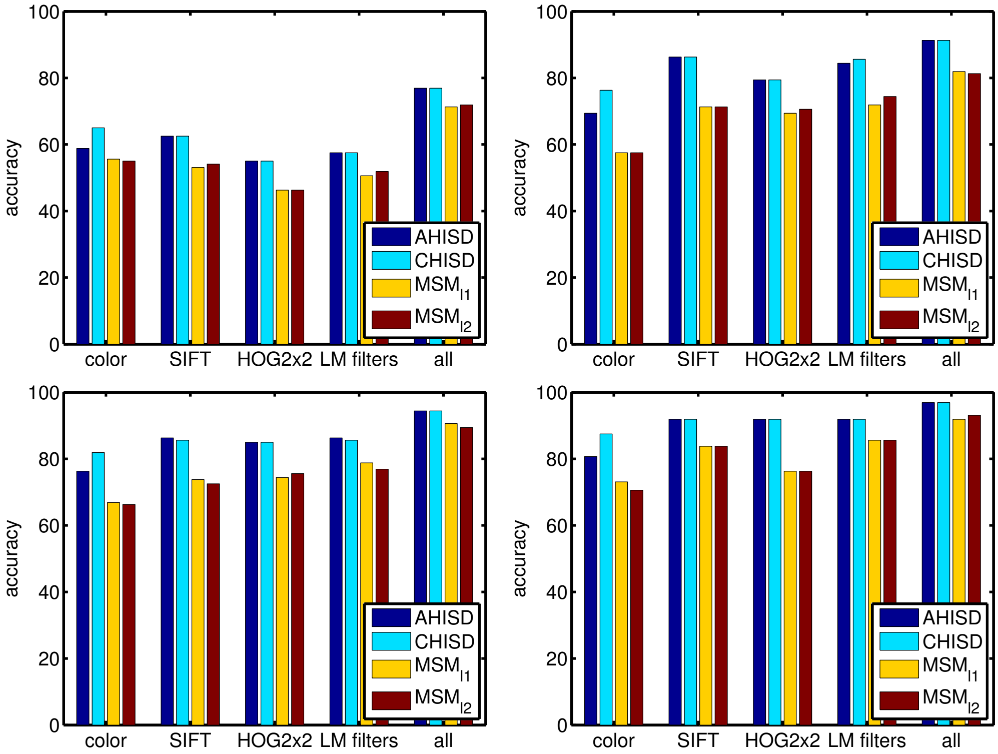


Figure 7.8: *Accuracies obtained for the data measured for the UBO2014 database and its extension when using VLADs, disjoint reference and query sets of* 5 *(upper left),* 10 *(upper right),* 15 *(lower left) and* 20 *(lower right) randomly drawn images taken under different view-light configurations.*

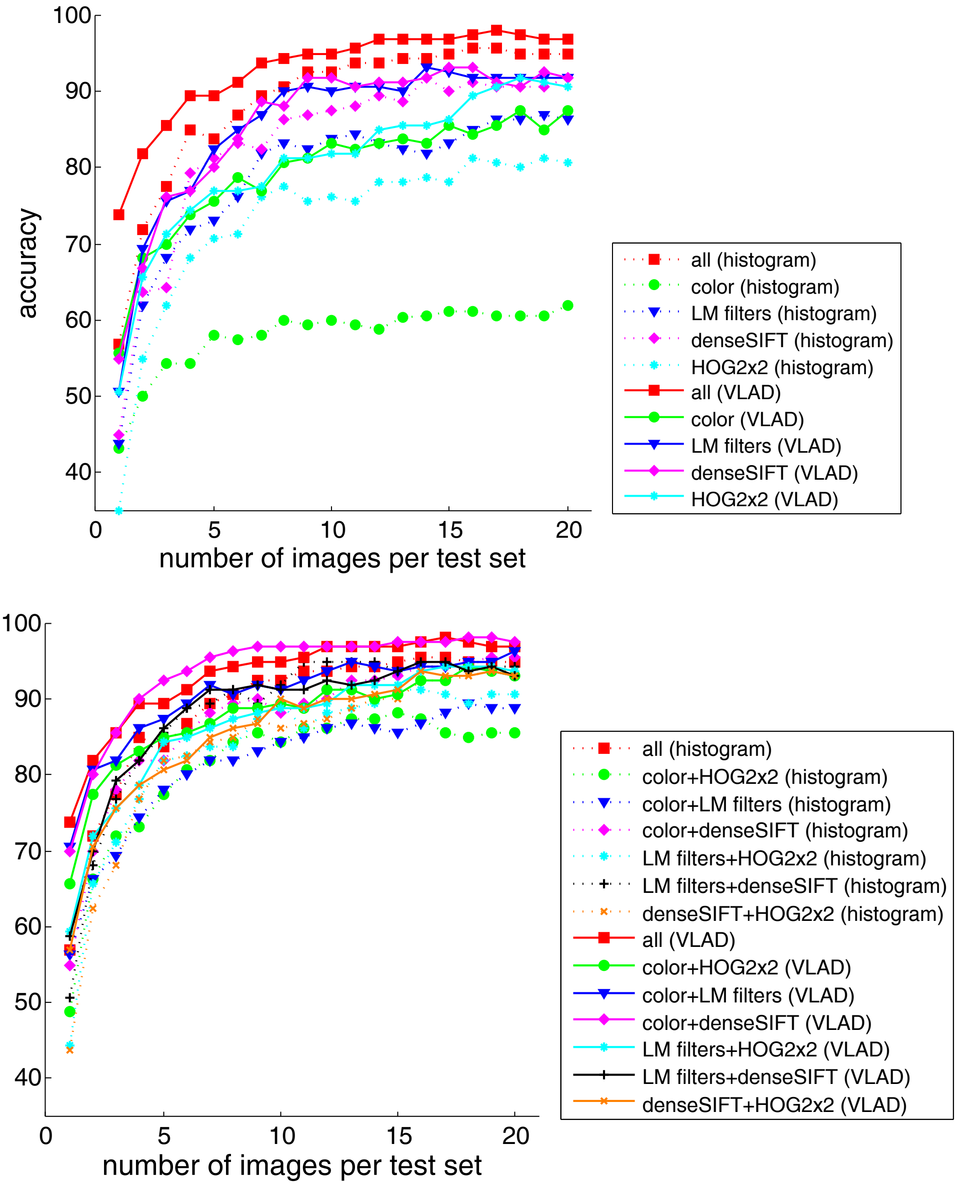


Figure 7.9: *Accuracies obtained for using sets of 20 view-light combinations for the reference sets and an increasing number of view-light combinations for the query materials (for the data measured for the UBO2014 database and its extension). As expected, the accuracy increases when using larger query sets.*

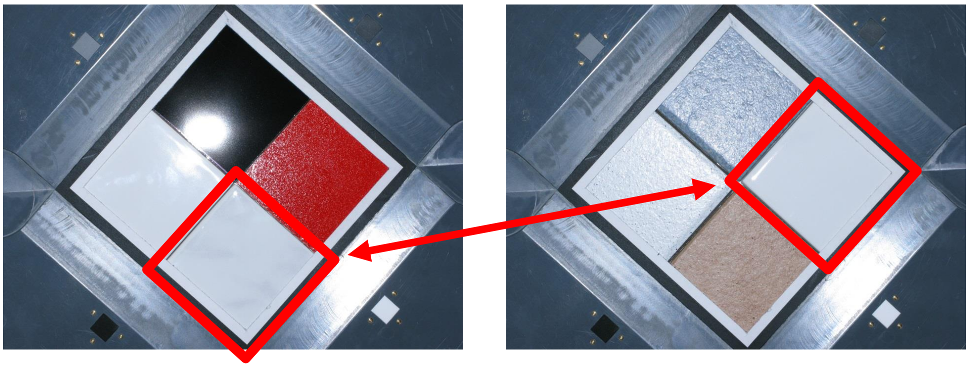


Figure 7.10: *Example showing misclassified materials. The estimated material and the ground truth material indeed look rather similar, and it is even hard for the human eye to distinguish them from most view-light configurations.*

CHAPTER 8

MATERIAL RECOGNITION UNDER NATURAL ILLUMINATION

The approach introduced in Chapter 7 focuses on material recognition in a lab environment with controlled illumination and possibly measurements of several view-light configurations, which is particularly interesting in scenarios with complex acquisition devices that allow measuring a multitude of view-light configurations. While such scenarios might be encountered in industrial environments e.g. where materials have to be sorted within supply chains, material recognition becomes significantly more challenging when leaving the lab conditions as it is required when analyzing materials in our natural environments. The main reasons for this can be identified in the complex interplay of material properties, the wide range of possible illuminations and varying surface geometries of the considered objects. In addition, there are usually only single images depicting materials under a single viewpoint and a single illumination configuration. This complicates the task even more. But how can we approach the task of recognizing materials in single images acquired under natural illumination?

To cope with the richness in appearance variation found in real-world data under natural illumination, we propose to synthesize training data that capture these variations for material classification. This work has successfully undergone peer review (see [WGK14]). Using synthetic training data created from separately acquired material and illumination characteristics allows to overcome the problems of existing material databases which only include a tiny fraction of the possible real-world conditions under controlled lab environments. However, it is essential to utilize a representation for material appearance which preserves fine details in the reflectance behavior of the digitized materials. As BRDFs are not sufficient for many materials due to the lack of modeling mesoscopic effects, we present a high-quality BTF database with 22,801 densely measured view-light configurations including surface geometry measurements for each of the 84 measured material samples. This representation is used to generate a database of synthesized images depicting the materials under different view-light configurations with their characteristic surface geometry using image-based lighting to simulate the complexity of real-world scenarios. We demonstrate that our synthesized data allows classifying materials under complex real-world scenarios.

After introducing the contributions of our technique for material recognition under natural illumination (see Section 8.1), we discuss standard databases used for material recognition (see Section 8.2) and discuss the generation of synthetic data used in the scope of this chapter (see Section 8.3). This is followed by a description of the recognition scheme (see Section 8.4) and a detailed evaluation (see Section 8.5). A short summary and a discussion of the limitations to be investigated in future work conclude the chapter (see Section 8.6).

8.1 Motivation

Image-based scene understanding depends on different aspects such as the detection, localization and classification of objects. For these tasks, it is essential to consider characteristic object properties such as shape or appearance. While its shape tells us how to grasp a particular object, its material tells us how fragile, deformable, heavy, etc. it might be and hence, how we have to handle it. The understanding of the recognized surface material thus guides the interaction of humans with the corresponding object in daily life, and it also represents a key component regarding industrial applications. However, image-based recognition of materials in real-world environments is a challenging problem due to the huge impact of viewing and illumination conditions on material appearance. Therefore, training an appropriate classifier requires training data that representatively covers all these conditions as well as the intra-class variance of the materials.

So far, there have been two main approaches to generate suitable training data. One approach is to capture a single representative per material category under a multitude of different conditions, such as scale, illumination and viewpoint, in a controlled setting [DvGNK96, HCFE04, CHM05, LYG13] (see Table 8.1). However, the measured viewing and illumination configurations are rather coarse and hence not descriptive enough to capture the mesoscopic effects in material appearance, which consider the light interaction with material surface regions mapped to approximately one pixel, in an accurate way. In addition, the material samples are only measured under controlled illumination or lab environments which does not generalize to material appearance under complex real-world scenarios. As an alternative, the second category of methods uses images acquired under uncontrolled conditions. In [SRA09], images from an internet image database (Flickr) have been used. The advantage of this is that both the intra-class variance of materials and the environment conditions are sampled in a representative way. Unfortunately, the images have to be collected manually, and the materials appearing in the images have to be segmented and annotated. The necessary effort again severely limits the number of configurations that can be generated this way (see Table 8.1). More recent databases such as the OpenSurfaces dataset [BUSB13] and the Materials in Context Database [BUSB14] contain even more annotated surfaces from real-world consumer photographs. However, annotating and segmenting several thousands of images contained in these databases has to be performed by crowd sourcing which, in contrast, is rather expensive.

In this chapter, we instead make use of synthesized data which has already been explored for different applications (e.g. [EG08, PJW+11, TGZ08, SGS10, SFC+11, ON12, BM12, BM13]). In particular, separately acquired material characteristics and illumination conditions offer the possibility to create synthetic training data for material recognition that capture the variations of real-world data. This decoupling of the sampling of material from environment conditions allows us to overcome the limitations of existing material databases that contain only a few hundred configurations of viewing and lighting conditions per material category. For these synthetic images, perfect segmentations are directly available without the need for manual segmentation and annotation, and a huge number of them can be obtained easily and fully automatically. This approach requires the creation of realistic renderings, which accurately simulate the appearance of a material in a real-world scenario. In particular, the appearance of many daily life materials such as cloth, skin, etc. is determined by effects taking place on surface structures mapped to a size of approximately one pixel (e.g. scratches or fibers) such as subsurface scattering, interreflections, self-shadowing and self-occlusion. These effects cannot be modeled by standard Bidirectional Reflectance Distribution Function (BRDF) models, which are suitable especially for locally smooth surfaces such as plastic or metal as these fulfill the assumption of a homogeneous surface reflectance behavior. This was pointed out in [WHON97], where the concept of Apparent BRDFs (ABRDFs) has been introduced to take the above-mentioned effects into account. Bidirectional Texture Functions (BTFs) [DNvGK97] are a data-driven approach to efficiently capture and store ABRDFs and represent these mesoscopic effects. The results in [LF12], where training data has been synthesized based on BRDFs, support exactly this claim by showing that using BRDF materials for synthetic training data alone is not sufficient for materials exhibiting mesoscopic effects of surface reflectance and leads to classification results significantly worse than using real-world images. In contrast, our experiments indicate that using an appropriate representation of the reflectance behavior such as the BTF opens up the possibility of solely using synthesized training data for classification tasks. We demonstrate that the classification of real-world test data can be boosted significantly by using image-based lighting via environment maps [Deb98] instead of simple directional light sources. To achieve this, we generate synthesized training samples under a vast amount of different lighting conditions simulated by arbitrary HDR environment maps, which adequately represent the complexity of real-world materials and lighting.

For this purpose, we have acquired a database containing dense BTF measurements of 84 material samples. The samples can be grouped into 7 categories (i.e. 12 samples per class). Per BTF, all combinations of 151 view directions and 151 light directions have been measured which results in $22{,}801$ images per sample or a total of $7 \cdot 12 \cdot 22{,}801 > 1.9$M images respectively. The data of our measured database with directional illumination is used as input to generate the synthesized data. By acquiring a height map of each material sample via the structured light technique described in Chapter 3, we also include the complexity of the geometric structure of the different materials in the process of generating synthetic training images. While in fact an arbitrary number of configurations could easily be included in the synthesized database, we so far used 42 different viewpoints and 30 different illumination conditions per material sample.

In summary, the key contributions presented in this chapter are:

- a technique that allows the decoupling of the acquisition of material samples from the environment conditions by generating synthetic training samples,
- a publicly available novel BTF database of 7 material categories, each consisting of measurements of 12 different material samples, measured in a darkened lab environment with controlled illumination,
- a second, novel database containing data synthesized under natural illumination which represents a clear difference to other datasets that only use directional illumination or an additional single ambient illumination, and
- an evaluation which shows that these synthetic training samples can be used to classify materials in photographs under natural illumination conditions.

8.2 Databases for Material Recognition

In this section, we briefly review commonly used databases for material recognition and discuss their limitations (see Section 8.2.1). Subsequently, we discuss approaches that follow the recent trend of using synthetic training data in various applications (see Section 8.2.2).

8.2.1 Conventional Material Databases

Table 8.1 gives an overview of several different material databases. The CUReT database [DvGNK96] contains measurements of 61 material samples under 205 configurations with different viewing angles and different directional illumination conditions. This database has been extended in the scope of the KTH-TIPS database [HCFE04] in terms of varying the distance of the acquired sample to the camera, i.e. the scale of the considered textures, in addition to changing viewpoint and illumination angle. In both databases, however, only a single material instance is provided per class, and thus the intra-class variation of semantic material categories is not represented. Aiming for a generalization to classifying semantic material categories, the KTH-TIPS database has been extended by adding measurements of different samples of the same semantic material category and also considering ambient lighting in the KTH-TIPS2 database [CHM05]. However, taking only four samples per category still limits the representation of the intra-class variance of material categories observed in real-world scenarios. More recently, a spectral material database has been presented in [LYG13] for multi-spectral material recognition. However, the samples are imaged from only one single viewpoint. A common limitation of all these databases is the rather limited number of measurements, which are furthermore acquired in a lab environment. Hence, the influence of the complexity of real-world environment conditions is not taken into account, and, therefore, material recognition under natural illumination cannot be performed based on such training data.

Other databases are designed to capture the large intra-class variation in the appearance of materials in complex real-world scenarios. The Flickr Material Database (FMD) [SRA09] contains images that have been downloaded from `Flickr.com` and show different associated material samples under uncontrolled viewing and illumination conditions and compositions. Even larger collections are given by the OpenSurfaces dataset [BUSB13] or the Materials in Context Database (MINC) [BUSB14]. However, annotations and segmentations of the images of these collections require plenty of work in a time-consuming process and are typically obtained by costly crowdsourcing services such as Amazon Mechanical Turk (AMT) [BUSB13, BUSB14, CMK$^+$14]. In addition, while manual segmentations are available, these masks are not always accurate, leading to the inclusion of background appearance and problematic artifacts for material recognition. Obviously, the significantly more complex variations of material appearance encountered under natural illumination make material recognition much more challenging and only recognition rates far below the ones obtained for databases acquired under controlled lab conditions have been reached so far [LSAR10, SLRA13]. The main reason for this is that it is more complex to include the possibly encountered

Table 8.1: *Overview of different databases. Please note that the FMD considers different configurations of viewing and lighting conditions as well as different material samples for each individual image. Our databases are highlighted in red (*: in principle, an arbitrary number of configurations could be considered in the synthesis).*

	CURET [DvGNK96]	KTH-TIPS [HCFE04]	KTH-TIPS2 [CHM05]	MPI-VIPS [LF12]	spectral database [LYG13]	measured database	FMD [SRA09]	synthesized database
material samples	61	10	44	11	90	**84**	1,000	**84**
categories	61	10	11	11	8	**7**	10	**7**
samples per category	1	1	4	1	N.N.	**12**	100	**12**
illuminations	4 ... 55	3	4	4	150	**151**	100	**30***
illumination type	controlled	controlled	controlled & ambient	controlled & ambient	controlled	**controlled**	natural	**natural**
viewpoints	7	27	27	27	1	**151**	100	**42***
images per sample	205	81	108	108	150	**22,801**	1	**1,260***
total image number	12,505	810	4,752	1,188	13,500	**1,915,284**	1,000	**105,840***

variations on material appearance in the training data than for material recognition under controlled illumination data, where a smaller subset of training data might already be sufficient. Furthermore, a different approach has been presented with the Describable Textures Dataset (DTD) [CMK+14]. While the aforementioned databases establish classes for different material instances or more general semantic material categories, this database considers semantic material attributes as classes. This allows to represent materials in terms of how well they match the individual attributes.

8.2.2 Previous Approaches Based on Synthetic Training Data

The manual processes typically required to capture exemplars as well as to segment and annotate materials in images severely limit the number of images per material category in all of the above-mentioned databases. As an alternative, the potential of computer graphics has been investigated to introduce a new promising trend of using synthesized training data for several applications.

Recombination methods focus on some specific aspects present in real-world examples and recompose them to new examples as done in [EG08, PJW+11] to enlarge the available training data by recombining shape, appearance and background information for pedestrian detection. In [TGZ08], new virtual training images are synthesized via photometric stereo for texture classification. This way, fewer training images need to be acquired. In contrast, rendering techniques can be used to produce new examples based on an underlying model. For instance, pose estimation was facilitated using synthesized depth maps in [SFC+11]. In [SGS10], object detection based on 3D CAD models is investigated using viewpoint-dependent, non-photo-realistic renderings of the object contours to learn shape models in 3D, which then can be matched to 2D images showing the corresponding object. Furthermore, an evaluation of the commonly used image descriptors based on a photo-realistic virtual world has been carried out in [KTF11]. This virtual scenario represents a well-suited setting to analyze the effect of illumination and viewpoint changes. The methods in [BM12] and [BM13] use a renderer to synthesize shading images based on given depth maps and a spherical harmonic model for illumination for the estimation of shape, illumination and reflectance from input images. This way, a decoupling of albedo and illumination is reached. The decoupling of measured surface material and environmental lighting has also been addressed in [ON12], where shape and BRDF of objects have been jointly estimated under known illumination from synthetic data generated from different combinations of shapes, environment illuminations and BRDFs. In [WD04], geometric textons are rendered under different view-light configurations in order to estimate geometric texton labels used in a hybrid model for geometry and reflectance.

Recently, this trend has resulted in the development of the virtual MPI-VIPS database introduced in [LF12] (see Table 8.1). This database is based on using BRDFs to represent the light exchange on the surface of an object and does not rely on physical measurements but uses a texture map and material shaders of available rendering packages. Bump maps are used to simulate the local mesostructure of the material surface to improve the shading effects. The selection of shaders, viewpoints and illuminations used to render the materials is closely oriented on the KTH-TIPS2 database. The texture map does not capture intra-class variance and the approximate rendering models result in a less realistic depiction of some materials such as aluminum foil, which appear rather artificial, especially in complex light situations. The reason for this less realistic impression of the synthesized materials is that the complexity of the reflectance characteristics of the involved materials has not been adequately considered as e.g. mesoscopic effects contributing to the appearance of many materials such as textiles, bread or cork are not modeled. In addition, the shaders are not suitable to accurately reproduce the reflectance behavior of crumpled aluminium foil. The investigations in [LF12] therefore indicate that a training set based on the utilized virtual samples alone performs poorly for material classification and a mixture of real and rendered samples is necessary to get acceptable results. In contrast to these studies, we show that the synthesis approach to generate virtual samples matters. Our measured database covers intra-class variances better and includes significantly more viewing and lighting configurations than any of the other databases. These dense measurements are required for the realistic depiction of many materials with their characteristic traits in a virtual scene via BTFs to preserve the mesoscopic effects in the synthesized data.

8.3 Generation of Synthetic Training Data

In this section, we discuss the details of our database of measured BTF material samples and how this database is used to produce synthetic training images of these samples under a range of different viewing and illumination conditions.

8.3.1 BTF Material Database

Since we intend to create synthetic training images, it is necessary to digitize the material samples in such a way that it becomes possible to reproduce the material appearance under nearly arbitrary viewing and lighting conditions. Though a wide range of material descriptions exists, image-based BTFs have proven to be a representation which is suitable for a wide range of materials, as already discussed in Section 8.1. Since their introduction in [DvGNK96], the technology has advanced considerably, and today devices for the practical acquisition of BTFs at high angular and spatial resolutions are available. A detailed survey on devices for reflectance acquisition is given in e.g. [HF13, SSW$^+$14, WK15a, WdBKK15, WLGK16]. In contrast to the small number of representative images acquired for the other databases listed in Table 8.1, these setups allow to acquire tens of thousands of images. Those images are taken in a lab environment and, hence, are not directly applicable for typical real-world scenarios. However, this much larger number of viewing and lighting conditions offers the possibility to render high-quality images of the materials under nearly arbitrary viewing and lighting conditions, where material traits are still accurately preserved. For a recent survey on BTFs, we refer to [FH09].

Our measured database is formed by 7 semantic classes which are relevant for analyzing indoor scenarios (see Figure 8.1). To sample the intra-class variances, each of these 7 material categories of our database contains measurements of 12 different material instances. These instances share some common characteristics of the corresponding category but also cover a large variability. With a total of 84 measured material instances, we provide more than the CUReT database, the KTH-TIPS database and the KTH-TIPS2 database (see Table 8.1). For each of the materials, we have measured a full BTF with $22{,}801$ HDR images (bidirectional sampling of 151 viewing and 151 lighting directions) of a 5cm $\times$ 5cm patch with a spatial resolution of 512×512 texels. Thus, our database contains more than 1.9

million images. Additionally, for each sample, a height map has been acquired via structured light. This helps to reduce compression artifacts and allows to render realistic silhouettes. We employed a reference geometry to evaluate the RMS error between the reconstruction and the ground truth geometry which was approximately 25μm as already mentioned in Chapter 3. The acquisition of both geometry and BTF of a material sample was achieved on a fully automatic basis in approximately 3 hours, and up to 4 samples can be acquired simultaneously. In particular, there is no need for manual annotation and segmentation which is not feasible for large image collections. Our database is available at `http://cg.cs.uni-bonn.de/en/projects/btfdbb/download/ubo2014/`.

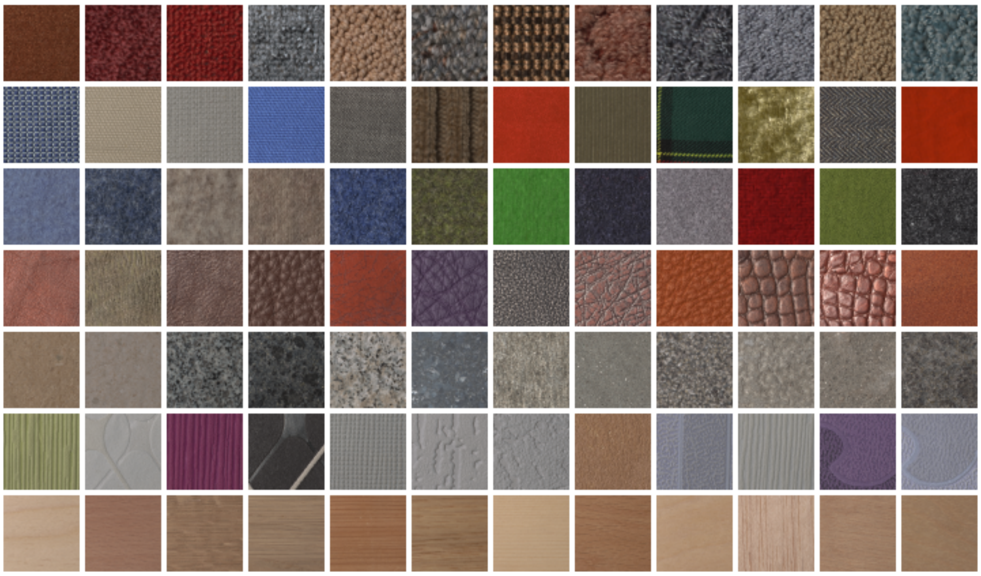


Figure 8.1: *Representative images for the material samples in the 7 categories.*

8.3.2 Synthesis of Novel Training Images

Once the materials have been measured, it is in principle possible to render images showing the materials under nearly arbitrary viewing and lighting conditions. In order to train a material classifier, we have to decide for which conditions we synthesize the training images, and we need a technique to synthesize a sufficiently large number of images efficiently. Additionally, the material representation used to produce the renderings needs to be capable of accurately depicting the traits in material appearance. In the synthesis process (see Figure 8.2), the measured geometry and BTF of a considered material sample are rendered under different illumination conditions simulated by environment maps, which is a standard technique in computer graphics (see e.g. [Deb98]). Furthermore, utilizing the measured geometry allows compensating parallax effects. The latter would otherwise be induced by surface regions which significantly protrude from the modeled reference surface and result in a blurring of the surface details. We followed the technique in [RSK13] which is based on the reprojection of the BTF onto the geometry. The result remains a BTF parameterized over the respective non-planar reference geometry and not a Spatially Varying BRDF, as the reflectance functions still remain data-driven ABRDFs. Hence, effects like interreflections, self-shadowing, etc. can still be reproduced. For geometric details not contained in the reference geometry, the major parallaxes are removed by the reprojection and the remaining disparities do not significantly affect the appearance of the synthesized material.

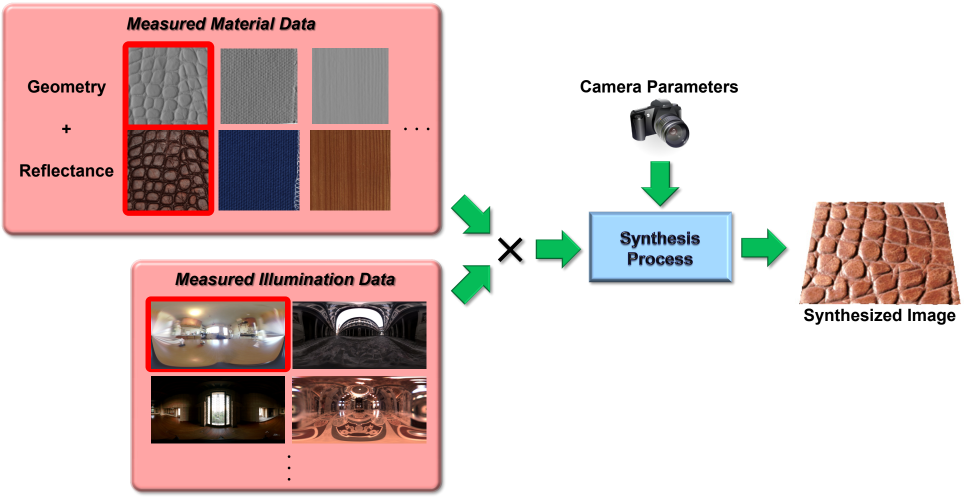


Figure 8.2: *Synthesis of representative training data: The full Cartesian product of material data (corresponding geometry and reflectance) and environment lighting (environments taken from [Deb13]) can easily be rendered by using a virtual camera with specified extrinsic and intrinsic parameters. The illustrated output image is generated using the material and illumination configuration highlighted in red.*

In the rendering process, the exitant radiance $L_r(\mathbf{x}, \omega_o)$ is calculated for each surface point $\mathbf{x}$ via the image-based relighting equation

$$L_r(\mathbf{x}, \omega_o) = \int_{\Omega} \mathrm{BTF}(\mathbf{x}, \omega_i, \omega_o)\, L_i(\omega_i)\, V(\mathbf{x}, \omega_i)\, \mathrm{d}\,\omega_i, \qquad (8.1)$$

where ω_i and ω_o represent the incoming and outgoing light direction. $L_i(\omega_i)$ denotes the radiance distribution in the environment map over the spherical domain

Ω. The visibility function $V(\mathbf{x}, \omega_i)$ represents a binary indicator function describing if the environment map is visible from surface point $\mathbf{x}$ in the direction ω_i. Due to the enormous number of images that we want to synthesize, the use of an efficient rendering technique is mandatory. Therefore, we decided to additionally use an OpenGL-based renderer to generate our database. To simulate the HDR environment in this renderer, we approximate it in a way similar to the work in [BARA06] with 128 directional light sources, distributed representatively over the environment via a relaxation algorithm. In this case, the equation for the evaluation of the exitant radiance $L_r(\mathbf{x}, \omega_o)$ reduces to

$$L_r(\mathbf{x}, \omega_o) = \sum_{\omega_i \in \mathcal{L}} \mathrm{BTF}(\mathbf{x}, \omega_i, \omega_o)\; L_i(\mathbf{x}, \omega_i)\; V(\mathbf{x}, \omega_i), \tag{8.2}$$

where $V(\mathbf{x}, \omega_i)$ represents a shadowing term computed via shadow mapping [RSC87] and $\mathcal{L}$ denotes the set of light source directions, i.e. the ω_i represent the directions towards the utilized directional light sources. That way, it becomes possible to render the images with a double resolution full-scene anti-aliasing at a resolution of $1{,}280 \times 960$ pixels in about 2s on a GPU, including the computation of the 128 shadow-maps necessary to compute $V(\mathbf{x}, \omega_i)$. Figure 8.3 illustrates the considerable variations in material appearance captured in the synthesized data due to changes in the illumination and viewing conditions.

For every combination of material sample and environment map, we then generated training images, depicting a planar material sample under a range of 21 different rotations of the material sample

$$(\theta, \varphi) \in \{0^\circ, 22.5^\circ, 45^\circ\} \times \{-67.5^\circ, -45^\circ, -22.5^\circ, 0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ\}$$

and in two different distances to also consider the scale-induced variations in appearance of the materials. To increase the variance captured by our dataset further, we also use 6 rotated versions of each of the used environment maps. As a consequence, we obtain $1{,}260$ images per material sample (see Table 8.1). Though we only used planar samples in the scope of this chapter, the BTFs could, in principle, also be rendered on arbitrary geometry to further increase the space of sampled conditions.

8.4 Recognition Scheme

Figure 8.4 illustrates our recognition scheme which shows similarities to our technique presented in the scope of Chapter 7. To make the current chapter self-containing, we briefly discuss the relevant components also in the scope of this chapter.

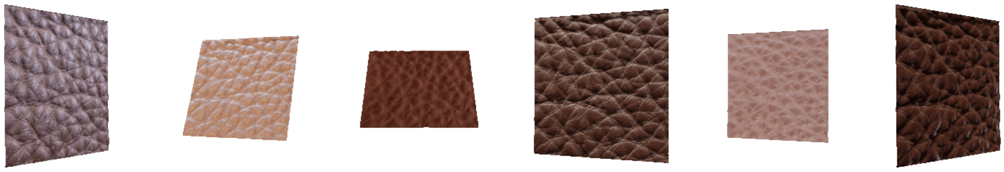


Figure 8.3: *Examples for synthesized images of the same material sample that demonstrate the large variation in material appearance under different viewing and illumination conditions.*

In order to capture different aspects of material appearance, we use densely sampled 3×3 color patches and densely sampled SIFT descriptors. Both of these descriptor types represent standard descriptors (e.g. [LSAR10, SLRA13]). Although the color of a material varies depending on the conditions of the environment and the viewpoint, it still contains valuable information as the variance of the color of a certain material sample under natural illumination is typically limited. Using dense SIFT features has become a popular choice in scene, object and material recognition [BZM06, ZMLS07, LSAR10, LF12, SLRA13] as well as these features capture the local spatial and directional distribution of image gradients and provide robustness to variations in illumination and viewpoint. In our system, these features are extracted on multiple scales ($s \in \{1, 2, 4, 6, 8\}$). Both descriptor types are extracted on a regular grid with a spacing of 5 pixels as in [LSAR10].

Once features have been extracted, an appropriate representation of the content of the image regions with the respective materials has to be computed for each type of descriptor. For this purpose, we first generate a dictionary of visual words for the individual feature types by k-means clustering [Ste57, Llo57, Llo82, Mac67] of the respective descriptors extracted from the images in the training set. As already mentioned in Chapter 7, this allows us to represent the single images either by histograms as used in standard bag-of-words (BOW) approaches or by more sophisticated representations such as Fisher vectors [PD07] or vectors of locally aggregated descriptors (VLADs) [JDSP10] which have shown to yield superior performance when compared to standard BOW approaches. Hence, we choose VLADs to describe the content of the masked regions. This means that all the local descriptors $\mathbf{x}_i$ in an image are first assigned to their nearest neighbor $\mathbf{c}_j$ with $j = 1, \ldots, k$ in the corresponding dictionary with k visual words for each feature type. Subsequently, the entries in the VLAD descriptor are formed by accumulating the differences $\mathbf{x}_i - \mathbf{c}_j$ of the local descriptors and their assigned visual words according to

$$\mathbf{v}_j = \sum_{\{\mathbf{x}_i | \mathrm{NN}(\mathbf{x}_i) = \mathbf{c}_j\}} \mathbf{x}_i - \mathbf{c}_j. \tag{8.3}$$

The final descriptor is built via the concatenation $\mathbf{v} = \left[\mathbf{v}_1^T, \ldots, \mathbf{v}_k^T\right]^T$. However, the dimensionality of this representation is rather high-dimensional ($d \cdot k$). Here, d represents the dimensionality of the local descriptors (e.g. $d = 128$ for SIFT) and k the number of words in the dictionary. We utilize PCA and take the 250 most relevant components of the PCA space per descriptor type for the training data. The VLAD representations of the test set are projected into this space.

The final classification task can be performed by using standard classifiers, such as the nearest neighbor classifier, random forests [Bre01] or support vector machines [Vap95]. The latter have already been successfully applied in the domain of material recognition [VZ09, CHM05, SLRA13]. Since an SVM with RBF kernel outperformed the nearest neighbor classifier or random forests in our experiments, we only report the numbers for the SVM, where the regularization parameter and the kernel parameter are estimated based on the training data using grid search.

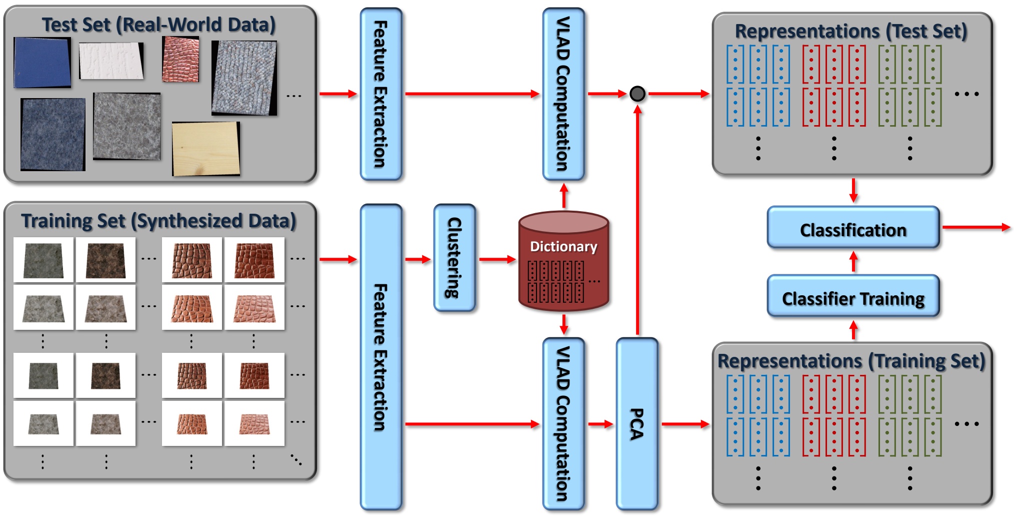


Figure 8.4: *Recognition scheme: Based on the descriptors extracted from the synthetic training data (where the masks for the presence of materials are automatically given) we calculate a dictionary via k-means clustering [Ste57, Llo57, Llo82, Mac67]. This dictionary is used to encode the descriptors per masked region via VLADs [JDSP10]. Then, a dimensionality reduction of these VLADs is performed via PCA which is followed by an SVM-based classification.*

8.5 Experimental Results

In the scope of our experiments, we focus on the question whether real-world materials can be classified using synthesized training data. For this purpose, we first validate our recognition scheme on standard material databases (see Section 8.5.1). In the next step, we perform a detailed evaluation of the use of our synthesized training data for material recognition (see Section 8.5.2) which is followed by a comparison to the use of other synthesized datasets (see Section 8.5.3). After this, we analyze the potential of our synthesized training data to recognize materials in internet photos (see Section 8.5.4). In order to obtain the VLAD representation of the individual feature types, we use dictionaries with 150 visual words for the color descriptor and 250 visual words for the SIFT descriptor in our experiments similar to [LSAR10, SLRA13].

8.5.1 Validation of Recognition Scheme on Commonly Used Material Databases

With accuracies of 99.11% and 99.25% on the CUReT database and the KTH-TIPS database respectively, our system is on par with recent state-of-the-art approaches as listed in [TVG12] which achieve accuracies of around 99%.

8.5.2 Analysis of Using Synthetic Training Data

Our main experiments target material recognition under everyday illumination conditions. For this reason, we acquired photographs of the samples of the 7 classes considering arbitrarily chosen poses of the camera with respect to the material samples for the test set $\mathcal{T}_{te,1}$. Different illumination conditions are taken into account by placing the material samples into different environments: a room with natural illumination, a room with a mix of natural illumination and neon lamps, a room with neon lamps and two darkened room scenarios with a rather directional illumination. In each of the 5 scenarios, each material sample is photographed twice using different viewpoints, which results in a test set of 840 images. Based on this test set, we evaluate whether our synthesized training data (both pathtraced and OpenGL-based) can be used to train a robust classifier. Additionally, we perform an evaluation of considering natural illumination vs. considering directional illumination as present in the measurement data. This will indicate what can be gained from the training data synthesized under natural illumination. The results are summarized in Table 8.2.

Comparison of Measured vs. Pathtraced Training Data (Directional Illumination): In a first step, we considered training the classifier using training data with illumination via point light sources. We randomly selected 50 images per material sample from the measured data resulting in a training set $\mathcal{T}_{tr,m}$ of $4{,}200$ images. Using this training set, we obtain a classification accuracy of 58.92% on $\mathcal{T}_{te,1}$. To support our assumption that virtual images are of a similar quality as their real-world counterparts, we generated a virtual duplicate of the utilized measurement device using the pathtracer implementation in [Jak10]. Using this virtual setup, we produce synthetic training data following Figure 8.2 for exactly the same viewing and illumination configurations as given in $\mathcal{T}_{tr,m}$. In this case, we illuminate the respective material samples using point light sources as given in the real device instead of using environment maps. The resulting classification accuracy of 60.48% closely matches the accuracy obtained for the real-world measurement data.

Comparison of Measured vs. Pathtraced Training Data (Natural Illumination): Here, we analyzed the effect of considering more complex illumination as encountered in typical real-world scenarios for the training. We captured all the 84 material samples under two representative room environments and an outside environment in a courtyard, and from two different viewpoints which results in a training set of 504 images. Based on this training set, where we expect the camera settings (viewpoints with respect to material samples, white-balancing, etc.) to be close to the ones used for the test set, we obtain a recognition accuracy of 75.83%. In order to synthetically simulate this scenario, we captured light probes of the three environments and used them to generate training data under more typical real-world lighting but under the same viewpoints as present in $\mathcal{T}_{tr,m}$. This leads to a training set of $12{,}600$ images for which we obtain an accuracy of 68.21%.

Obviously, there is a clear benefit of using representative environments in the generation of the easy-to-produce synthetic training data in comparison to the illumination via point light sources as present in $\mathcal{T}_{tr,m}$. In addition, the characteristic material traits seem to be preserved sufficiently within our synthesized data to allow the classification of real materials. However, we recognize a difference in performance between using the training set of $12{,}600$ images synthesized under environment lighting and the use of the 504 photos taken in the respective environments. This might be due to the noise introduced by the pathtracer with only 32spp (samples per pixel) which influences the descriptors as well as due to not perfectly matching the assumptions of far field illumination and the neglection of emitting surfaces. The reason for only taking 32spp is that data generation using a pathtracer takes a lot of time, especially if different environment lighting and different scales are desired. Rendering the $4{,}200$ images for the virtual measurement device (under one single environment map) for instance already takes about two days using 32spp with our implementation based on Mitsuba on a Intel Xeon CPU E5-2690v2 workstation (32 cores, 3GHz). We also did not perform a white-balancing of the data under the environment lighting which might influence both descriptor types. Furthermore, the acquisition conditions (view conditions, camera characteristics) of both $\mathcal{T}_{te,1}$ and the 504 real-world training images were similar.

Comparison of Measured vs. Rasterized Training Data (Natural Illumination): As a consequence of the slow rendering via a pathtracer, we used our OpenGL-based synthesis procedure to generate the huge amount of images in our synthesized database. As a training set, we consider a random subset of 600 different viewing and illumination conditions from this synthesized data for each of the classes resulting in $4{,}200$ images. In this scenario, our classifier yields a classification accuracy of 72.74% which again significantly outperforms the accuracy of 58.92% obtained when using $4{,}200$ photos acquired during the measurement of the samples in a lab with controlled illumination for the training. It even almost reaches the accuracy of 75.83% achieved in the experiment mentioned before. This might be due to the fact that we do not encounter the problem of noise induced by the pathtracing approach when using the OpenGL-based synthesis as well as due to better matching the viewpoint conditions in $\mathcal{T}_{te,1}$ by accounting for multiple scales.

Furthermore, we analyzed the impact of using different numbers of the OpenGL-synthesized images for the training. The accuracy increases with an increasing size of the training data, which is to be expected, as larger training sets cover a larger variance of the utilized viewing and illumination conditions (Table 8.2).

Comparison of Per-Class Accuracies: There seems to be a trend that in particular the samples of the categories fabric, felt, leather and stone can be categorized more reliably when using the synthesized training data (OpenGL-based) with natural illumination in comparison to measurement data with directional illumination (improvements of around 22% (fabric), 10% (felt), 30% (leather), 35% (stone) and less overfitting for the remaining categories). This agrees with our motivation for this study as we expect the samples of these classes to have more variance in appearance under the different illumination conditions due to their deeper mesostructure and their surface reflectance behavior.

Classifier Generalization to Unseen Material Samples in Different Environments Based on Synthesized Data: For each of the classes, we draw a random subset of 600 images with different viewing and illumination conditions from the

Table 8.2: *Classification of the manually acquired photos in $\mathcal{T}_{te,1}$ using different training sets. (*: pathtraced using Mitsuba renderer [Jak10]; **: OpenGL-based synthesis using 4 environment maps available from [Deb13] and 1 environment map available at our department).*

training set	**illumination type**	**type of training data**	**performance on $\mathcal{T}_{te,1}$**
4,200 images from measurement	directional	real-world	58.92%
4,200 synthesized images (pathtraced*) using the same viewing and lighting conditions as present in measurement	directional	synthetic	60.48%
12,600 synthesized images (pathtraced*) using the same viewing conditions as present in the measurement data but under 3 measured environments	natural	synthetic	68.21%
504 photos acquired in 3 measured environments	natural	real-world	75.83%
525 synthesized images (OpenGL-based**)	natural	synthetic	62.74%
1,050 synthesized images (OpenGL-based**)	natural	synthetic	65.71%
2,100 synthesized images (OpenGL-based**)	natural	synthetic	68.69%
4,200 synthesized images (OpenGL-based**)	natural	synthetic	72.74%

complete synthetic training data. We split the material samples of the 7 classes into disjoint training and test sets by using 8 material samples observed under 4 different environments for the training set and the remaining 4 samples rendered under the fifth environment map as the test set. The resulting accuracy of 62.29% indicates the ability of our classifier to generalize to unseen material samples and illumination conditions. Using more material samples per category and more environment maps would probably lead to an increase in accuracy.

8.5.3 Using Our Synthesized Database vs. Using Previous Synthesized Training Data for Classifier Training

A comparison to other approaches using synthesized data, such as [LF12], is not directly possible. While the material shaders and the selected illumination conditions utilized for the generation of synthetic data in [LF12] are chosen to correspond up to some degree to the conditions during the acquisition of the KTH-TIPS2 database, our synthesized data considers different material categories which we expect to be more relevant for scene analysis because of their presence in offices, buildings and streets. Our data does not focus on the controlled illumination conditions in a lab environment and, instead, approaches the more complex real-world conditions in arbitrary environments. The only class we directly share with the MPI-VIPS and the KTH-TIPS2 databases is wood. In order to analyze the difference of using shaders with the reproduction of the illumination conditions present in the test dataset and our data synthesized from several real-world wood samples under more complex illumination conditions, we first train a wood-vs-rest classifier on the synthesized MPI-VIPS database and perform a classification on the KTH-TIPS2 database. Here, 61.57% of the images in the wood category of the KTH-TIPS2 database are recognized correctly. In contrast, we also perform an experiment where we have replaced the wood images of the MPI-VIPS database in the training set with images taken from our OpenGL-synthesized data with environment illumination for the class wood. Even though the illumination conditions in our data are rather different in comparison to the ones present during the acquisition of the KTH-TIPS2 database, we obtain a correct recognition of 76.16% for the images with the label wood of the KTH-TIPS2 database, which represents a significant improvement. This clearly demonstrates the benefit of using accurately digitized materials for material recognition from synthesized data and taking the intra-class variances into account.

8.5.4 Material Recognition in Internet Photos

For each of our 7 material categories, we downloaded 20 images and performed a manual segmentation on each image. Then, the masked material regions form the set $\mathcal{T}_{te,20}$. Taking a subset of 15 images per class from $\mathcal{T}_{te,20}$ gives another test set $\mathcal{T}_{te,15}$. Using our aforementioned training data of $4{,}200$ images synthesized using OpenGL and considering environment illumination gives accuracies of 65.71% ($\mathcal{T}_{te,15}$) and 62.86% ($\mathcal{T}_{te,20}$). In comparison, using $\mathcal{T}_{tr,m}$ for the training results in an accuracy of only 56.19% for $\mathcal{T}_{te,15}$ and 56.43% for $\mathcal{T}_{te,20}$.

In addition, training the classifier on 5 of the images per class not included in $\mathcal{T}_{te,15}$ gives an accuracy of 41.90% on $\mathcal{T}_{te,15}$. The influence of adding synthesized data to this training set on the accuracy obtained for $\mathcal{T}_{te,15}$ as well as a summary of the other results in this paragraph are shown in Table 8.3. Taking more training data with a larger variance of the utilized illumination conditions and the utilized viewpoints leads to an increasing performance. This clearly demonstrates the power of using synthesized materials for practical applications.

Except for the category leather, we also used the samples present in the CUReT database to represent the categories. For each category, we selected 92 images equally distributed over the material samples contributing to the classes (carpet:

samples 18,19; fabric: samples 2,3,7,22,29,42,44,46; felt: sample 1; stone: samples 10,11,17,30,33,34,36,37,41,49,50; wallpaper: samples 12,31,38; wood: samples 54,56). In this experiment, we obtained accuracies of 41.11% ($\mathcal{T}_{te,15}$) and 36.67% ($\mathcal{T}_{te,20}$) which indicates a bad generalization of the CUReT database to natural illumination, varying viewing conditions and intra-class variances. Furthermore, the image quality is rather low for the CUReT database.

Table 8.3: *Classification of internet images ($\mathcal{T}_{te,15}$ and $\mathcal{T}_{te,20}$) using different training sets. (*: OpenGL based synthesis using* 5 *environment maps available from [Deb13]; †: category leather is not covered in the CUReT database).*

training set	**illumination type**	**type of training data**	$\mathcal{T}_{te,15}$ 15 internet images	$\mathcal{T}_{te,20}$ 20 internet images
CUReT images †	directional	real-world	41.11%	36.67%
4,200 images from measurement	directional	real-world	56.19%	56.43%
4,200 synthesized images*	natural	synthetic	65.71%	62.86%
internet images	natural	real-world	41.90%	–
internet images+ 4,200 synthesized images*	natural	mixed	66.67%	–
internet images+ 16,800 synthesized images*	natural	mixed	72.38%	–

8.6 Conclusions

In this chapter, we have presented an approach to create synthetic training samples for material recognition. This way, it is possible to decouple the acquisition of the material samples from the acquisition of the illumination conditions under which the material is observed. In addition, using synthesized data overcomes the necessity for time-consuming manual acquisition, annotation and segmentation of images. To evaluate our approach, we acquired a database of BTFs, containing 7 classes with 12 samples each, from which the training data is generated. Our evaluation demonstrates that synthetic training data offers new potentials regarding material recognition in complex real-world scenarios with natural illumination and clearly outperforms the alternative of taking images from the measurement of the material samples under controlled illumination conditions as training data. Therefore, our approach represents a significant step towards classifying materials in everyday environments which makes our approach valuable for many applications. Further work might be spent on extending the database by additional material classes in the future. In addition, increasing the number of viewing and lighting conditions considered in the synthesized database could be a future objective as well.

Part IV

Discussion

CHAPTER 9

Efficient, Automatic Selection of Required Acquisition Techniques

As mentioned in Chapter 2, there is a huge multitude of acquisition techniques where each of the individual methods is typically tailored to only a certain type of surface reflectance behavior. However, objects might consist of several different surface parts with different types of surface reflectance and appropriate acquisition methods might not be known a-priori. After a brief recapitulation of the motivation behind the techniques developed in the scope of this thesis (see Section 9.1), we discuss possibilities to combine the 3D shape acquisition techniques presented in Chapter 3, Chapter 4 and Chapter 5 in order to handle objects with diffuse, glossy and mirroring surface parts (see Section 9.2).

Furthermore, based on the availability of appropriate, material-specific acquisition techniques as discussed in Part II and the availability of reliable tools for material recognition as discussed in Part III, an efficient pipeline for both geometry and reflectance acquisition can be realized. This is required if there is no prior knowledge available regarding the surface materials of the object to be measured (see Section 9.3). The latter concept has successfully undergone peer review (see [WK15b]).

9.1 Motivation

The goal of accurately capturing details in surface geometry and reflectance behavior has led to a huge number of different acquisition methods and respective setups. However, current state-of-the-art acquisition procedures are designed regarding the expected reflectance behavior of an object to be digitized as already discussed in Part II.

In the domain of reflectance acquisition, it is well-known that smooth, homogeneous materials can be represented well with analytical BRDF models. These typically solely depend on the direction of the incoming light and the view direction. Spatially varying BRDFs additionally allow the modeling of spatial variations in surface reflectance behavior. However, materials exhibiting mesoscopic effects of light exchange on surface structures imaged to a size of approximately one pixel cannot be modeled by using simple BRDF models or spatially varying BRDFs. For such materials, current state-of-the-art techniques acquire data-driven BTFs which consider the spatial material variations in addition to the view direction and the direction of the incoming light.

In a similar way, 3D reconstruction techniques typically also depend on some basic assumptions about material reflectance. Many of the methods such as most multi-view stereo techniques and photometric stereo techniques are based on assuming Lambertian reflectance behavior (see Section 2.3.1). Some more sophisticated extensions also allow to consider the wider range of opaque surfaces (see Section 2.3.2). Furthermore, structured light systems are well-suited for the geometry acquisition of objects with a reflectance behavior ranging from diffuse to even specular as long as a sufficient diffuse reflectance component is present. In contrast, other reconstruction techniques are tailored to mirroring surfaces (see Section 2.3.3), translucent surfaces (see Section 2.3.4) or refractive surfaces (see Section 2.3.5). All the aforementioned geometry reconstruction techniques consider only a fraction of the possible surface materials and are – by design – not capable of handling arbitrary surface reflectance.

Without a-priori knowledge about the material properties of the considered object or material sample, a naïve acquisition strategy would be to apply several different material-specific techniques and merge their results. However, in many cases this is highly inefficient regarding acquisition time, and hardware components are stressed unnecessarily as many of the taken images do not have an influence on the final reconstruction and, hence, have to be neglected. For more efficient geometry and reflectance acquisition procedures in case of missing information about the material properties, it is therefore desirable to automatically select only the appropriate techniques instead of applying several different methods that might not necessarily be needed.

Consequently, this chapter introduces two novel concepts for efficient acquisition pipelines:

- A novel theoretical concept that allows to combine our 3D shape acquisition techniques presented in Part II to an automatic, efficient acquisition system (see Section 9.2) where the local consistencies obtained with the individually applied methods are explored in order to select the locally most suitable acquisition method. The resulting conceptual framework is based on establishing consistencies using cues that are individual to different acquisition techniques that, in turn, are designed for different types of surface reflectance behavior. Once such different consistencies have been extracted, they can be integrated into the same efficient optimization framework to get a 3D reconstruction for the considered object.

- A novel automatic, smart geometry and reflectance acquisition framework (see Section 9.3), in which a highly accurate material recognition step is used to select the required techniques for both geometry and reflectance acquisition.

9.2 3D Shape Acquisition of Objects With Unknown Surface Reflectance

Today's geometry acquisition methods still meet their limitations when prior knowledge about the surface reflectance behavior of the considered object is not available, as they are typically designed to be appropriate for only a rather small range of possible material types. However, the material type is often not known a-priori and, consequently, the automatic selection of appropriate acquisition methods represents an important prerequisite. Therefore, when several 3D geometry acquisition techniques are available, where each of them is capable of handling different types of surface reflectance, one important question is whether these different methods can be combined in a single automatic acquisition pipeline. Such a pipeline would not only allow the acquisition of objects consisting of a single a-priori unknown material but also allow the acquisition of objects with heterogeneous surface reflectance behavior where different acquisition techniques have to be used for the individual parts made of different materials. The following discussion will focus on a conceptual analysis regarding possibilities to combine the shape acquisition techniques introduced in the scope of Chapter 3, Chapter 4 and Chapter 5. The methods developed in the scope of these chapters are:

- a technique to acquire the geometry of Lambertian objects via photometric stereo as discussed in one of the experiments in Section 5.5,
- techniques to acquire the geometry of objects with a reflectance behavior consisting of both a diffuse and a specular component (see Chapter 3 for a purely structured light based approach and Chapter 4 for a combination of a structured light system with a multi-view Helmholtz stereopsis technique), and
- a technique to acquire the geometry of mirroring objects (see Chapter 5).

As described in Chapter 4, the geometry information reconstructed via structured light systems such as the one introduced in the scope of Chapter 3 can be transformed to a corresponding volumetric consistency and, hence, to a volumetric representation. Furthermore, one key observation can be identified in the fact that our reconstruction methods in Chapter 4 and Chapter 5 are based on information in a volumetric representation as well, and the respective optimization frameworks used for both methods are rather similar. Therefore, the task of combining these methods can be approached by combining the individual consistency measures derived per point in the volume using each of the acquisition techniques. In more detail, the structured light information and the Helmholtz normal consistency information measured as described in Chapter 4 have to be complemented with the normal consistency information for mirroring objects discussed in Chapter 5. This results in a modification of the energy functional with respect to the ones used for the individual approaches.

Another possibility to combine methods for Lambertian and mirroring objects could be the fusion of our geometry acquisition techniques based on multi-view normal field integration introduced in Chapter 5. In fact, combining the normal information obtained from photometric stereo measurements as briefly discussed in Section 5.5 and the normal information derived for mirroring objects following the technique presented in Section 5.4 leads to a purely normal-based geometry reconstruction process. For instance, the object depicted in Figure 9.1 exhibits heterogeneous surface reflectance characteristics as it has both rather matte and almost ideally mirroring parts. Based on the acquisition procedures discussed in the scope of this thesis, the first key observation is the fact that, for diffuse surface parts, normal hypotheses obtained via photometric stereo or Helmholtz stereopsis exhibit a rather small variance at points close to the surface whereas normal hypotheses derived via the technique for mirroring surfaces will typically not be consistent at the true surface. In contrast, on mirroring parts, the observation will be different as the normal hypotheses derived via the technique for mirroring surfaces will typically show a highly consistent, dominant mode whereas the normal hypotheses obtained via photometric stereo or Helmholtz stereopsis will not form a highly-dominant cluster but show a large variance at the true surface. While both of these types of normal information are separately used to acquire the geometry of Lambertian and mirroring surfaces, their combination to handle inhomogeneous objects can be approached by simply combining the local normal consistencies obtained for the individual methods in the volume. In order to make a robust reconstruction possible, an appropriate combination of the individual consistencies per volumetric point could be performed by e.g. taking only the normal information with the higher consistency. Consequently, it is not only possible to identify the local material type and the appropriate acquisition technique based on the consistencies obtained for the individual methods, but also to select the local surface normal computed via the respective technique.

Subsequently, based on the local normal estimates and their respective consistency estimate resulting after the combination of the individual methods, the same min-cut-based optimization framework as discussed in Chapter 5 can directly be applied for the final reconstruction. Therefore, in addition to providing information regarding material properties inferred from the consistencies, combining our methods would allow the reconstruction of the full 3D shape for the range from diffuse to mirroring objects with complex surface geometry.

Other investigations also derive normal information for specular objects (e.g. [BS03, CGS06, YIX07, NWR08, FCM+08, BHB11]), transparent objects (e.g. [MK05, KS05, MK07, YIX07, KS08, YWT+11]) or translucent objects (e.g. [CGS06, DMZP14, IMMY14]), and this normal information could probably also be integrated into our geometry acquisition framework. Particularly the integration of volumetric normal information for translucent objects or glass objects and the fusion of the respective consistencies might even improve the range of material types for which the shape of the respective objects can be reliably acquired.

Unfortunately, the aforementioned concepts of combining the individual methods require full measurements of the involved objects with all the different acquisition techniques that are available for the individual material types as illustrated in the naïve, automatic acquisition pipeline in Figure 1.3. As a result, the total acquisition time t_Σ can be computed as the sum of the acquisition times t_i with $i = 1, \ldots, N$ needed for the N individual measurements according to

$$t_\Sigma = \sum_{i=1}^{N} t_i. \tag{9.1}$$

Furthermore, hardware components might also be stressed unnecessarily as some of the measurements are not required for the final reconstruction and some of the computational effort is not required as well. This renders such an acquisition strategy rather impractical. As we would favor an efficient, automatic acquisition process only involving the required effort regarding measurement time and hardware usage, our investigations in the scope of the next section will focus on the information that can be inferred visually from images of a material surface and show whether this information can be used for a more efficient acquisition.

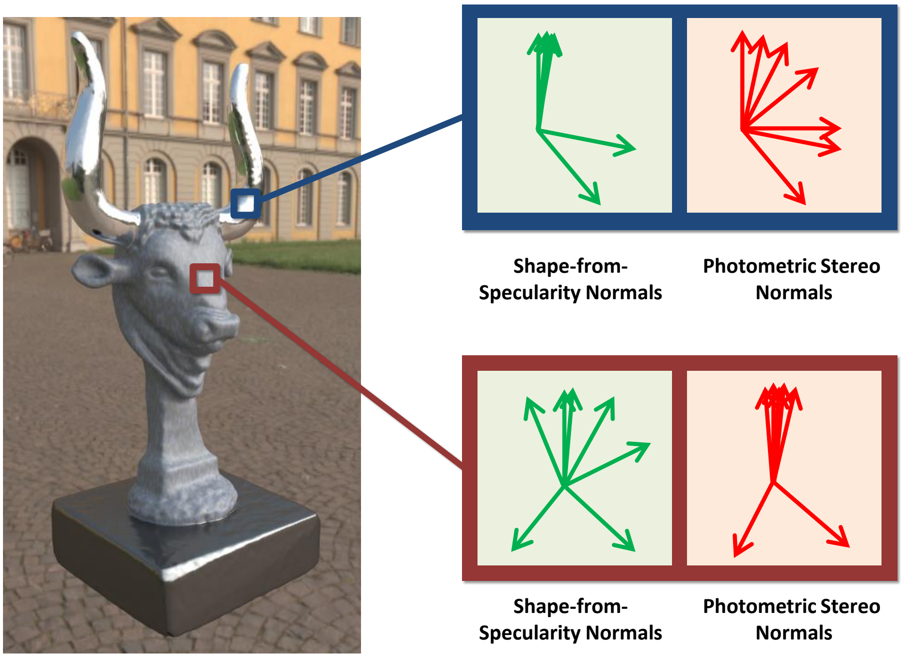


Figure 9.1: *Normal consistency for different acquisition methods: The shown object consists of rather matte and mirroring parts. When acquired with the methods developed in this thesis, normal hypotheses obtained via photometric stereo or Helmholtz stereopsis (red) show a small variance at surface points on the rather matte parts, whereas normal hypotheses derived via the technique for mirroring surfaces (green) will not be consistent. For points on the mirroring parts, the normal hypotheses derived via the technique for mirroring surfaces (green) form a dominant mode with a small variance, whereas the normal hypotheses obtained via photometric stereo or Helmholtz stereopsis (red) are not consistent on the surface.*

9.3 Efficient Automatic Acquisition Based on an Initial Material Recognition

As pointed out in [Rui13], cost-efficient measurements of complex objects or materials will probably necessitate the utilization of some kind of prior information. However, an open question is whether typical assumptions such as analytical BRDF models, smoothness, low-rank, etc. provide the best priors or whether data-driven priors represent a superior approach. In the scope of this section, we propose a complete pipeline for an efficient, automatic acquisition of geometry and reflectance. In particular, we demonstrate how the components developed in this thesis can be composed to an efficient, fully automatic acquisition framework where the required acquisition techniques are selected based on a prior material recognition (see Figure 9.2). In order to let the involved material recognition system act as an automatic assistance system which guides the subsequent acquisition process in such a way that only the required acquisition methods are involved, our approach relies on using a-priori information in the form of a database of material measurements with additional annotations regarding the methods that have to be chosen for the acquisition of geometry and reflectance. The key idea is the classification of a measured material based on only a rather small set of photos and, depending on the annotations of the closest match in the database, corresponding methods can easily be determined. For an almost mirroring metal, for instance, a shape-from-specularity approach could be used for geometry reconstruction and a BRDF measurement could be started to capture surface reflectance properties. If the considered material sample is classified as a material with strong mesoscopic effects as e.g. given in leather, a method appropriate for glossy materials such as our technique described in Chapter 4 could be used in combination with a BTF acquisition. The crucial prerequisite of such a strategy is the availability of a reliable material recognition framework, which has led to our investigations in Part III, and the availability of a database of materials with annotations regarding their appropriate acquisition processes as well as the availability of acquisition methods that are suitable for the different surface materials. Hence, one core component of our framework can be identified in a material database which contains images of a multitude of material samples taken under different viewing and lighting conditions, which are expected to be met during the acquisition with standard devices as analyzed in [SSW+14]. In our experiments, we used the material measurements performed for the BTF database introduced in Chapter 8 and additional material measurements resulting in data for 160 different material samples similar to our material recognition experiments described in Chapter 7 with additional annotations regarding the acquisition routines to be selected. In agreement with our results in Chapter 7, our material recognition has an accuracy of more than 97% based on only 20 view-light configurations per material and the few misclassified query material samples indeed have such a high visual similarity to the respective ground truth materials that they can hardly be distinguished by human observers as illustrated in Figure 9.3. In turn, this suggests that we might still take the stored parameters for a subsequent acquisition or reconstruction respectively due to the similarity of the materials. This can easily be verified by taking a closer look at the annotations regarding the acquisition methods to be chosen. In fact, 100% of the annotations of the query materials have been correctly estimated based on the recognition technique described in Chapter 7, i.e. our system allows a highly reliable recommendation regarding appropriate acquisition and reconstruction methods. Consequently, an appropriate acquisition process can be performed based on these suggested techniques for geometry and reflectance acquisition.

So far, our technique makes a significantly more efficient acquisition of geometry and reflectance behavior of material samples possible. Current limitations can be identified in the fact that its current version is only capable of reliably acquiring surfaces made of a single material type and not of several material types simultaneously. To handle such cases as well, our technique would have to be extended by an additional segmentation of the input images according to the respective materials. Per local material component, our technique might then be applied in a similar way. Furthermore, our technique for material recognition cannot handle non-planar objects yet as such objects require a more complex approach which considers the local variations of material appearance induced by the surface geometry and illumination conditions. While this still represents a challenge to be solved in the future, one possible strategy could be to consider local surface patches separately in the material recognition step and involve a patch-based application of the geometry acquisition techniques.

In order to get an impression about the amount of time that can be saved with an efficient automatic acquisition system based on a smart selection of acquisition techniques for the involved materials, we provide a short discussion with respect to the acquisition times. Our structured light based geometry acquisition technique described in Chapter 3 requires images that depict observed stripe patterns as projected onto the object surface by the involved projectors. The respective acquisition typically takes approximately 1.5 hours when using our highly parallel setup presented in [SWRK11] and in the range from approximately 1.5 hours to 3.0 hours in our turntable-based acquisition device introduced in [SSWK13, SSW+14]. In contrast, our geometry acquisition method introduced in Chapter 4 requires a full structured light measurement as well as a full reflectance measurement as input. Therefore, the acquisition time is significantly larger as the reflectance acquisition usually requires approximately 4 hours to 10 hours depending on the complexity of the surface materials. Furthermore, the geometry acquisition using our method presented in Chapter 5 takes approximately 1.5 hours to 2.0 hours when using illuminations from two to three different positions of the involved screens. When analyzing the required processing times, the higher complexity of our consistency-based techniques in Chapter 4 and Chapter 5 becomes apparent in processing times of approximately 8 hours to 12 hours on a workstation with two Intel Xeon 5645 CPUs with 2.4 GHz or approximately 4 hours to 5 hours on a workstation with two Intel Xeon E$5 - 2650$ CPUs with 2.0 GHz. In contrast, the triangulations required for the structured light approach in Chapter 3 and the subsequent Poisson reconstruction are carried out in approximately 1.5 hours.

Furthermore, the acquisition time as well as the processing time required for the reflectance reconstruction also strongly depend on the respective material. For materials following simple homogeneous BRDF models, only four parameters have to be measured, which might be done in only a few minutes. In contrast, materials with e.g. a spatially varying reflectance behavior require the measurement of six parameters in order to get an SVBRDF or BTF representation. In addition, the sampling density of the involved view-light configurations that need to be measured also needs to be taken into account. For dense measurements of several thousands of view-light configurations as used in e.g. [SWRK11, SSWK13, SSW$^+$14], even fine details of mesoscopic reflectance can be appropriately captured. However, even when acquiring the computationally demanding BTF, the acquisition and processing times might vary significantly depending on the complexity of the involved materials, which influences the acquisition parameters such as the number of exposures used during the measurement. This can be seen e.g. in the evaluation given in [Sch14], where the acquisition parameters and the processing parameters for different objects are given as listed in Figure 9.4, Figure 9.5 and in Figure 9.6.

As a result, the automatic selection of appropriate geometry and reflectance acquisition techniques with the respective postprocessing methods might save several hours.

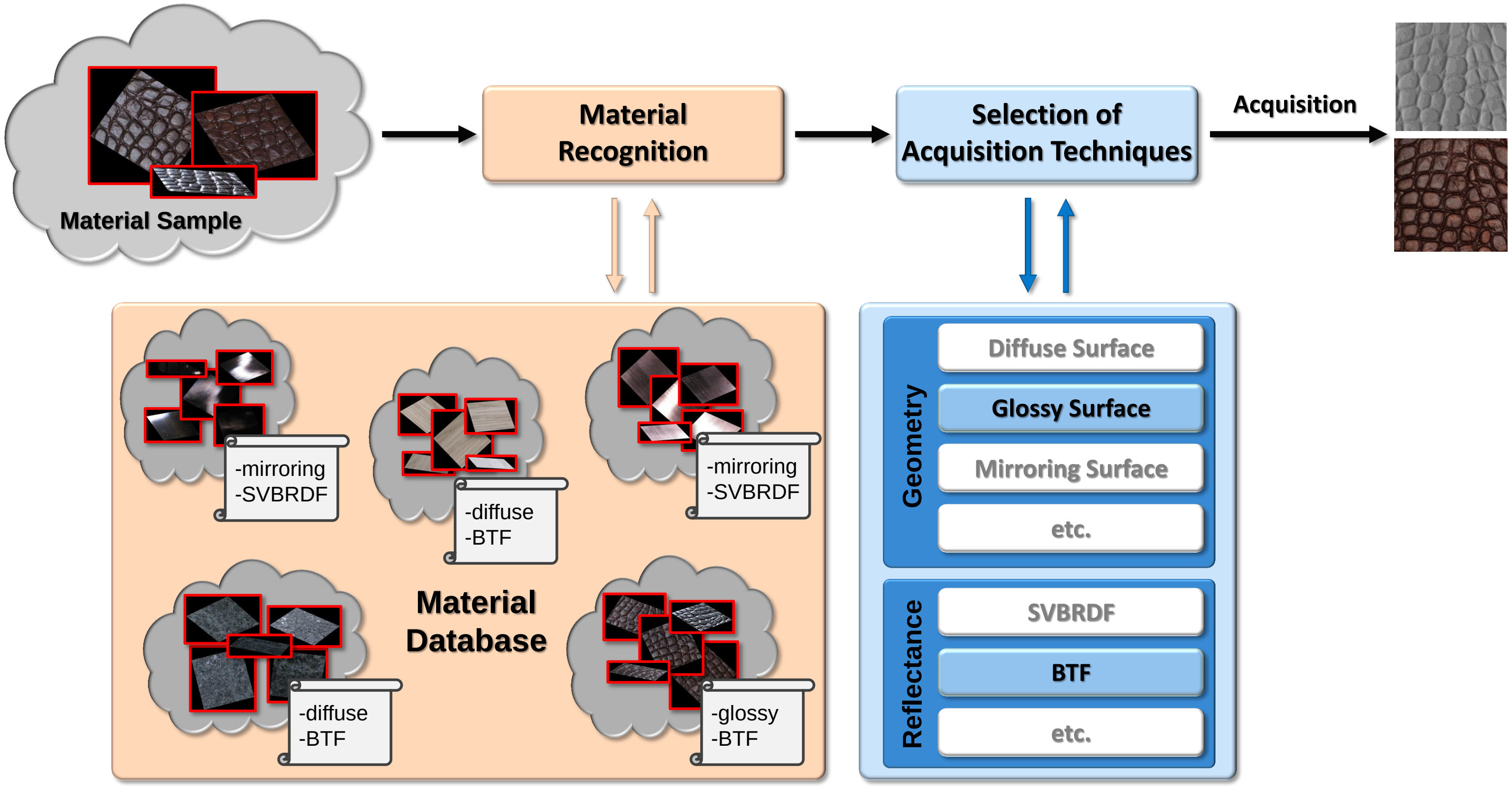


Figure 9.2: *Efficient, automatic acquisition pipeline. Based on a few input images of the material that should be acquired, an initial material recognition is used to find appropriate acquisition parameters in a database of reference materials with the respective annotations. Subsequently, these attached annotations for the closest material in the database can be used to guide the subsequent acquisition process by selecting the appropriate acquisition techniques. In the scope of our work, we focus on the acquisition of 3D geometry. However, the same pipeline also allows to select appropriate reflectance acquisition techniques.*

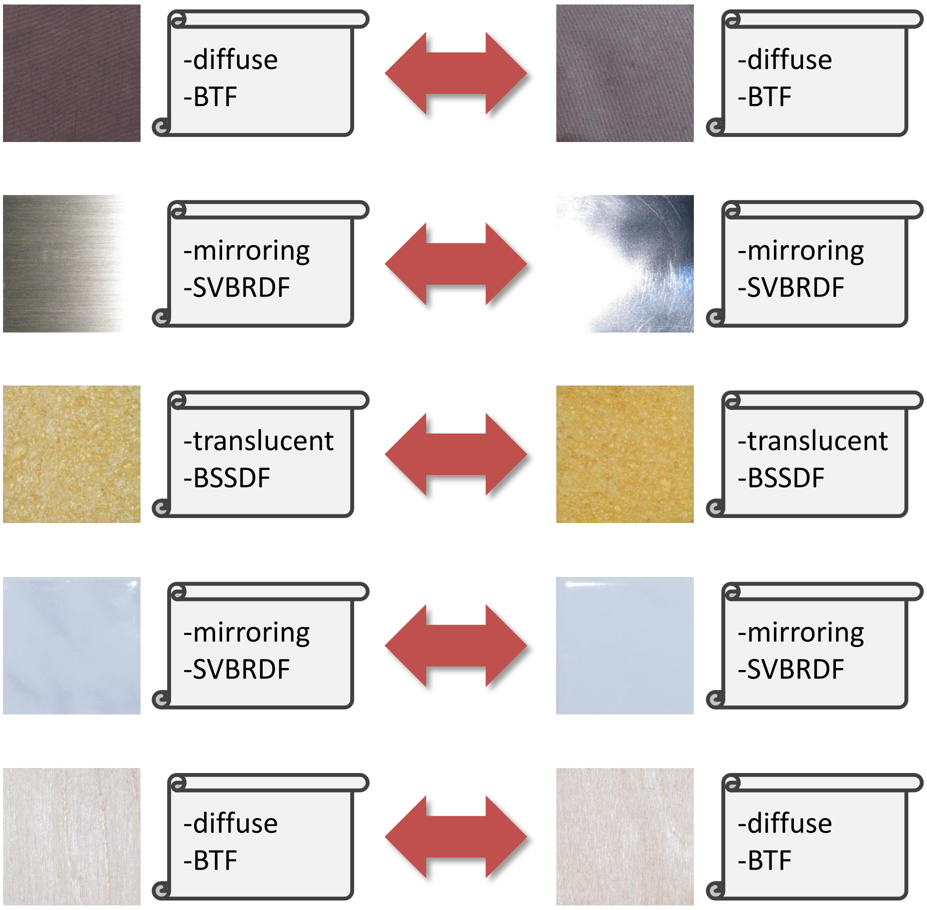
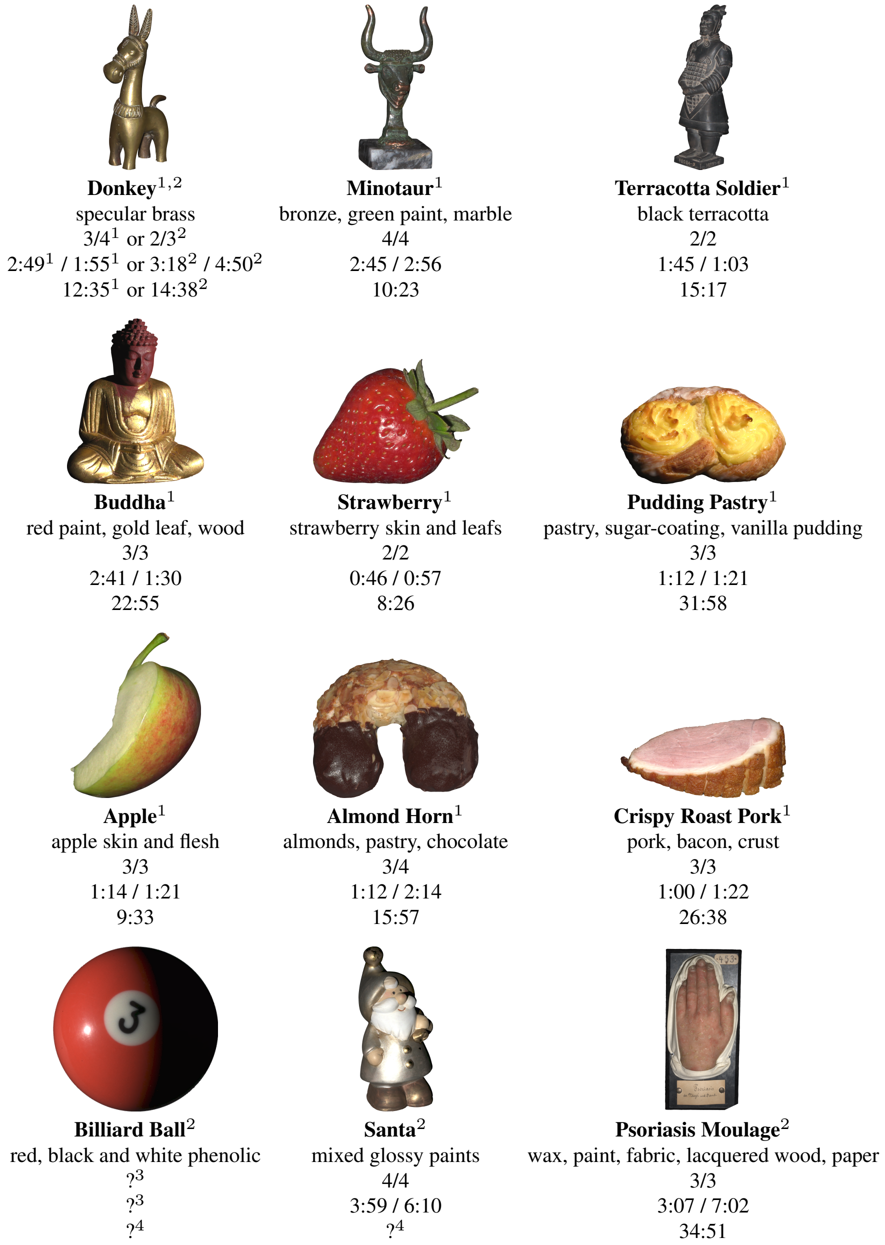


Figure 9.3: *Illustration of the incorrectly classified materials and their annotations: Material recognition has been performed based on* 20 *disjoint view-light configurations in the reference sets and the query sets and using color patches, SIFT descriptors, Gabor filters and HOG features. While for these few misclassifications the query materials (left) and their respective closest matching material in the database (right) still look rather similar, the annotations regarding the recommended acquisition systems are identical.*



Donkey[1,2]
specular brass
3/4[1] or 2/3[2]
2:49[1] / 1:55[1] or 3:18[2] / 4:50[2]
12:35[1] or 14:38[2]

Minotaur[1]
bronze, green paint, marble
4/4
2:45 / 2:56
10:23

Terracotta Soldier[1]
black terracotta
2/2
1:45 / 1:03
15:17

Buddha[1]
red paint, gold leaf, wood
3/3
2:41 / 1:30
22:55

Strawberry[1]
strawberry skin and leafs
2/2
0:46 / 0:57
8:26

Pudding Pastry[1]
pastry, sugar-coating, vanilla pudding
3/3
1:12 / 1:21
31:58

Apple[1]
apple skin and flesh
3/3
1:14 / 1:21
9:33

Almond Horn[1]
almonds, pastry, chocolate
3/4
1:12 / 2:14
15:57

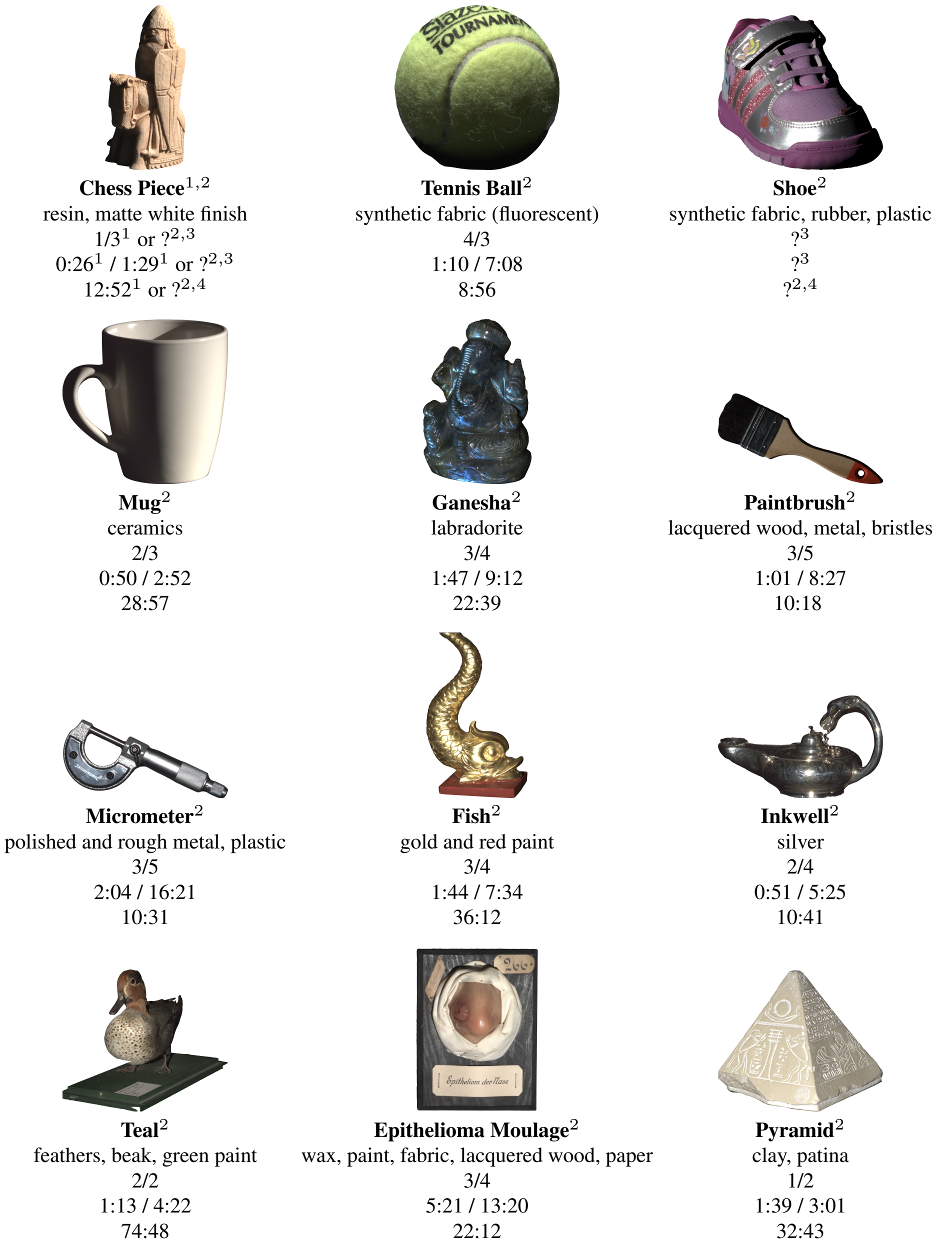
Crispy Roast Pork[1]
pork, bacon, crust
3/3
1:00 / 1:22
26:38

Billiard Ball[2]
red, black and white phenolic
?[3]
?[3]
?[4]

Santa[2]
mixed glossy paints
4/4
3:59 / 6:10
?[4]

Psoriasis Moulage[2]
wax, paint, fabric, lacquered wood, paper
3/3
3:07 / 7:02
34:51

Figure 9.4: *Objects digitized in the scope of [Sch14]: Listed are the respective apparent materials of the individual objects, the number of used exposures (geometry acquisition / reflectance acquisition), the acquisition times (geometry acquisition / reflectance acquisition) as well as the total processing times ([1]Measured in the Dome 1 device, [2]Measured in the Dome 2 device, [3]Original measurement data damaged, [4]No information available due to data loss).*



Chess Piece[1,2]
resin, matte white finish
1/3[1] or ?[2,3]
0:26[1] / 1:29[1] or ?[2,3]
12:52[1] or ?[2,4]

Tennis Ball[2]
synthetic fabric (fluorescent)
4/3
1:10 / 7:08
8:56

Shoe[2]
synthetic fabric, rubber, plastic
?[3]
?[3]
?[2,4]

Mug[2]
ceramics
2/3
0:50 / 2:52
28:57

Ganesha[2]
labradorite
3/4
1:47 / 9:12
22:39

Paintbrush[2]
lacquered wood, metal, bristles
3/5
1:01 / 8:27
10:18

Micrometer[2]
polished and rough metal, plastic
3/5
2:04 / 16:21
10:31

Fish[2]
gold and red paint
3/4
1:44 / 7:34
36:12

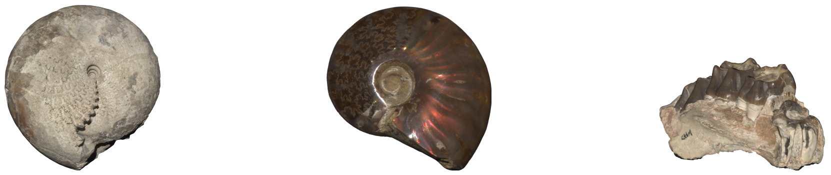
Inkwell[2]
silver
2/4
0:51 / 5:25
10:41

Teal[2]
feathers, beak, green paint
2/2
1:13 / 4:22
74:48

Epithelioma Moulage[2]
wax, paint, fabric, lacquered wood, paper
3/4
5:21 / 13:20
22:12

Pyramid[2]
clay, patina
1/2
1:39 / 3:01
32:43

Figure 9.5: *Objects digitized in the scope of [Sch14]: Listed are the respective apparent materials of the individual objects, the number of used exposures (geometry acquisition / reflectance acquisition), the acquisition times (geometry acquisition / reflectance acquisition) as well as the total processing times ([1]Measured in the Dome 1 device, [2]Measured in the Dome 2 device, [3]Original measurement data damaged, [4]No information available due to data loss).*



Ammonite 1[2]
fossil
2/3
1:04 / 6:24
19:36

Ammonite 2[2]
ammolite
3/5
3:20 / 20:13
9:28

Rhinoceros Teeth[2]
fossil
2/3
2:19 / 4:57
18:10

Figure 9.6: *Objects digitized in the scope of [Sch14]: Listed are the respective apparent materials of the individual objects, the number of used exposures (geometry acquisition / reflectance acquisition), the acquisition times (geometry acquisition / reflectance acquisition) as well as the total processing times ([1]Measured in the Dome 1 device, [2]Measured in the Dome 2 device, [3]Original measurement data damaged, [4]No information available due to data loss).*

CHAPTER 10

CONCLUSIONS

In this final chapter, we provide a summary of the contributions achieved in the scope of this thesis in Section 10.1. This includes an overview on the presented novel, robust 3D shape acquisition techniques for different surface materials in the full range from diffuse objects to mirroring objects (see Section 10.1.1), our novel material recognition frameworks for scenarios with controlled and natural illumination (see Section 10.1.2) and the novel concepts regarding efficient, automatic acquisition processes (see Section 10.1.3). Subsequently, we discuss the limitations of each technique and potential future research in Section 10.2.

10.1 Summary

In this section, we summarize the novel techniques that have been developed in the scope of this thesis in the context of 3D geometry acquisition (see Section 10.1.1), material recognition (see Section 10.1.2) as well as our novel concepts with respect to an efficient, automatic acquisition process based on our individual techniques (see Section 10.1.3).

10.1.1 Contributions Achieved in the Context of 3D Shape Acquisition

The contributions described in the chapters of Part II have been dedicated to the improvement of the state-of-the-art with respect to geometry acquisition techniques and handling different types of surface materials.

First, we have introduced a novel multi-camera, multi-projector super-resolution framework for structured light based geometry acquisition (see Chapter 3). This framework is capable of significantly increasing the density of the reconstructed point cloud in comparison to standard structured light systems. As it is typical for standard structured light systems, our framework is tailored to objects ranging from diffuse to even specular surface reflectance as long as a sufficient diffuse reflectance component is given. In this context, we observed that using high dynamic range imaging additionally improves the robustness of our structured light technique.

Secondly, we have designed a novel geometry acquisition framework in order to compensate the limitation of triangulation-based reconstruction techniques by using normal information in addition to the information derived with a structured light system (see Chapter 4). The method is capable of handling the range of opaque materials as the combination of a structured light system with Helmholtz normals is used. Both types of information are efficiently combined based on a novel efficient, volumetric surface reconstruction technique that uses the continuous min-cut on an adaptive grid structure and a subsequent refinement. In comparison to triangulation-based 3D reconstructions, our method allows a significant improvement in the accuracy of the reconstructed models. However, mirroring, translucent and transparent surfaces still cannot be handled by this approach.

Finally, we also introduced a novel geometry acquisition framework for the class of mirroring objects which cannot be reliably reconstructed using the aforementioned techniques (see Chapter 5). While still taking benefit from the efficient, volumetric surface reconstruction technique introduced in Chapter 4, the respective energy functional in the optimization is only based on normal information as the projector-based structured light system used in Chapter 3 and Chapter 4 cannot be used for the reconstruction of mirroring surfaces. As demonstrated in Chapter 5, our normal-based optimization framework has been proven to be highly accurate for the reconstruction of the full 3D shape of mirroring objects with even complex surface geometry including concavities and self-occlusions. In addition, our normal-based optimization framework represents a valuable approach for general normal-based 3D reconstruction approaches as any normal information that has been a-priori determined can serve as input. These normals can be inferred via different techniques that are appropriate for the individual types of surface reflectance behavior. In this regard, we have demonstrated that e.g. normals inferred via the standard photometric stereo technique can be used in our optimization framework to obtain a robust reconstruction for non-mirroring objects. Although in this example the considered object has a specular component in addition to the diffuse component and therefore violates the assumptions of the standard photometric stereo technique, a rather accurate model for the 3D geometry can be reconstructed.

All of our developed methods are designed in a systematic way that allows them to be combined in a single acquisition pipeline as discussed in Section 9.2. This would enable the automatic acquisition of the full 3D shape for objects with diffuse surface reflectance up to mirroring surface reflectance. Other normal-based information might, in addition, be easily integrated to even further increase the range of materials that can be acquired with the pipeline.

10.1.2 Contributions Achieved in the Context of Material Recognition

In this section, we will provide an overview regarding the contributions achieved in the context of material recognition and shortly hint on the typical scenarios for applying these techniques.

In the scope of the technical contributions with respect to material recognition discussed in Part III, we considered two scenarios. The first scenario, material recognition under controlled illumination, probably represents the typical scenario in industrial applications, where it is likely that material appearance might also easily be captured under several view-light configurations. Our approach presented in Chapter 7 has been demonstrated to be capable of reliably recognizing materials from only a few images measured under different view-light configurations. In contrast, the second scenario considers material recognition under natural illumination conditions. For this case, we presented a novel recognition scheme based on synthetic training data which circumvents the tedious manual work with respect to the acquisition of material samples under different view-light configurations, their respective annotation and their segmentation in the images (see Chapter 8). In particular, our method is designed to easily combine separately acquired material data and environment illumination data. Virtual but yet realistic material samples can be arbitrarily placed in synthetic environments and the viewing conditions can be specified arbitrarily as well.

To the best of our knowledge, both of our approaches represent a significant contribution in the respective research domain and might influence future work in this domain. In the scope of this thesis, we have investigated the impact of a reliable material recognition with respect to more intelligent and, hence, more efficient acquisition strategies in Part IV.

10.1.3 Contributions Achieved in the Context of Efficient, Automatic Acquisition Concepts

In the scope of this thesis, we have also presented two initial concepts towards a fully automatic 3D shape acquisition process for objects with a-priori unknown surface reflectance.

We have discussed a concept on how our geometry acquisition methods might be combined to handle the full range of diffuse to mirroring objects which can even be extended towards other types of surface reflectance such as existing for translucent or transparent objects. The key prerequisite of such methods for being integrated into the acquisition system can be identified in the normal fields that have to be estimated via appropriate techniques. Once the normals have been estimated, all the information can be fused using our highly efficient optimization framework based on the continuous min-cut formulation on an adaptive grid structure with a subsequent refinement step.

Furthermore, we have also pointed out the potential of an initial material recognition stage for a selection of appropriate acquisition techniques which is certainly of great importance regarding both geometry and reflectance acquisition (see Section 9.3).

10.2 Limitations and Future Work

After having summarized the achievements of this thesis, this section is focused on the discussion of limitations of our methods as well as future avenues of research to compensate them. We also provide an outlook to upcoming trends in the related domains.

10.2.1 Geometry Acquisition

While the techniques presented in Chapter 3, Chapter 4 and Chapter 5 have led to significant improvements with respect to the reconstruction of diffuse, opaque and mirroring objects, these methods are still not suitable for high-quality reconstructions for the full range of possible surface materials. In particular, our methods face limitations with respect to the acquisition of the geometry of translucent and transparent objects.

While several investigations explored shape-from-specularity techniques to acquire transparent objects (e.g. [YIX07, YWT$^+$11]) and translucent objects (e.g. [CGS06, GAVN11]), there are some critical aspects that might prevent reliable reconstructions for the respective surface materials. Handling translucent objects might e.g. be complex when using shape-from-specularity techniques such as our technique for the reconstruction of mirroring objects as the pattern information might get blurred at the transitions of the stripe patterns observed on the translucent object. As the resulting decoding of the structured light codewords will be less reliable, the reconstruction quality will decrease as well. To overcome this problem, it might be worth to consider different structured light encodings than Gray codes as used in the scope of this thesis. In this regard, the combination of conventional Gray codes and specific Gray codes with a larger minimal stripe width has been investigated in [GAVN11]. An alternative could be to apply the separation of direct and global components of the scene as presented in [NKGR06].

Finally, the proposed rather material-specific acquisition techniques rely on an a-priori known information regarding the surface material of the object to be acquired. Based on this information, a decision can be made on whether the acquisition method is appropriate or not. Usually, this information is part of the experience of e.g. a cultural heritage expert who selects the appropriate acquisition technique depending on the surface material of the object. However, in many cases it is desirable to consider fully automatic solutions for 3D reconstruction as e.g. in a supply chain, and typically there is no a-priori information about the surface material available as this might vary for each object. Similarly, different parts of an object could be made of different surface materials. Therefore, the respective object might have a heterogeneous surface reflectance with e.g. diffuse and mirroring parts. For this purpose, the appropriate combination of the individual techniques as discussed in Chapter 9 would be useful for an automatic acquisition. While we only have presented a theoretical concept for the combination of our individual methods, the practical implementation still has to be carried out in the future. In addition, suitable 3D reconstruction methods for objects made of materials with other types of surface reflectance such as transparent or translucent objects need to be integrated. While our considerations focused on geometry acquisition in lab environments where the illumination and viewing conditions can be controlled, there is an emerging interest in techniques for *in the wild* geometry acquisition that can ideally be performed based on sparse input data acquired via mobile devices such as mobile phones. Such conditions are particularly challenging for more complex surface reflectance behavior and adequate methods have to be addressed by future research. In recent years, a few methods have been proposed which focus on bridging the gap of photometric stereo methods so that they can also be applied in scenarios with natural illumination (see e.g. [ALFG12, MKSN12, YYT$^+$13, GYN13]).

10.2.2 Material Recognition

In the scope of our material recognition schemes, we currently use a combination of off-the-shelf color and texture descriptors. While these descriptors already help to achieve acceptable recognition rates, even more effort has to be spent on the development of material descriptors. An emerging trend that has gained particular attraction simultaneously to the developments of this thesis can be identified in learning descriptors for material recognition [SN13, LF14, CMK$^+$14, BUSB14] instead of relying on pre-defined techniques for extracting color descriptors or texture descriptors such as filterbanks [LM99a, LM01, Sch01, FA91, Gab46], Local Binary Patterns (LBPs) [OPM02], HOG descriptors [DT05], SIFT descriptors [Low99, Low04], SURF descriptors [BETvG08], basic image features [CG08, CG10] or sorted-random-projection descriptors [LFCK12]. This allows the identification of the material structures that are most important to achieve a reliable material recognition. In addition, attribute-based descriptions [CMK$^+$14] that consider attributes such as bumpy, checkered, dotted, fibrous, knitted, porous, smeared, sprinkled, stained, striped, woven, or zigzagged also represent an important direction of future research. Therefore, even more effort has to be spent on the development of attribute-based datasets such as the Describable Texture Dataset [CMK$^+$14] which consider variations in the appearance of attributes in a better way. In this context, it might also be worth to analyze the subjective perception criteria such as warm/cold, rough/soft, etc. in addition to semantic, visual features such as glossiness or roughness as these attributes guide the material selection process of designers as well as the material editing.

Considering the material reflectance in different spectra as e.g. the near-infrared or the ultra-violet spectral range might additionally contribute to a more robust material recognition, as the appearance of materials observed under these spectral ranges might be distinguished more easily than in the RGB channels. In this regard, material appearance in the near-infrared domain has already been considered in e.g. [SFS09], and in [LYG13], illumination by six different wavelengths is used to classify materials.

Furthermore, the widely used material databases presented in [DvGNK96, HCFE04, CHM05, BG09, SRA09] contain only a rather small number of images with respect to the variations in material appearance encountered in daily life, which is not sufficient for a generalization to other databases where materials are photographed under different viewing and illumination conditions and on different surface geometries. For this purpose, much larger material databases are required. Recent approaches towards larger databases include OpenSurfaces [BUSB13], which contains annotated surfaces from real-world consumer photographs, and our database of synthesized materials as described in Chapter 8. While several thousands of hours have been spent for the annotation and segmentation of the data for the OpenSurfaces database, our approach in Chapter 8 does not require manual annotation and segmentation for the training data. However, to handle the wide range of appearance variations of individual classes such as fabrics, leather, stone or wood, a much larger number of materials would have to be measured for each of the classes in comparison to the 12 material samples per class provided by our database. Furthermore, more material categories have to be included. For some rather specular materials, the generation of synthesized data might be based on the fitting of analytical BRDFs as the BTFs are not suitable to reproduce strong specularities. In addition, to allow material classification on complex surface geometries, it might be favorable to use a multitude of different surface geometries to generate synthesized data similar to the data mentioned in Chapter 8 and derive a novel, even better representation for materials. Obtaining such amounts of measured material data is currently still a time-consuming process as measurements in standard devices as e.g. the camera arrays in [SWRK11, SSW$^+$14] require an acquisition time of several hours per measurement. However, the trends for faster appearance acquisition devices will probably reduce the acquisition time significantly in the future and will make our approach more practical. Again it is worth mentioning that many of the respective schemes towards sparse reflectance acquisition (see e.g. [dBSHK14]) or multiplexed reflectance acquisition (see e.g. [dBSHK15]) only work for a certain range of materials. Furthermore, we have performed an investigation of linear models to even further reduce the number of view-light configurations measured during the acquisition in [dBWK15].

Having huge masses of data, i.e. many images, there is an additional need for efficient large-scale learning techniques that can train per-class models based on a high number of images in a reasonable time. Furthermore, when huge masses of tens of thousands of view-light configurations systematically acquired by devices such as the ones presented in [SWRK11, SSW$^+$14] are available, it might also be possible to identify view-light configurations that are most informative regarding material recognition. These might e.g. be useful for the calculation of better convex hull models for our technique described in Chapter 7. Additionally, these most important view-light configurations might provide important insights regarding the development of small, hand-held devices for material recognition. Obviously, only a very limited number of view-light configurations can be considered in the design of such a device, which makes the consideration of the most informative view-light configurations an important prerequisite. By simulating the acquisition process in a virtual device via the currently available rendering frameworks similar to our approach in Chapter 8, fundamental information might be inferred regarding an optimal configuration of the components involved in the measurement under which material characteristics can be successfully detected, e.g. by analyzing the classification/retrieval performance.

So far, we did not consider an automatic segmentation of materials within images. Certainly, this represents an important objective in the future as well, as it reduces the involved manual work significantly. In this context, more investigations towards color segmentation strategies with robustness with respect to shadows, highlights, and textures such as the one in [VBvdWV11] will have to be carried out.

The insights provided by this thesis in the context of material recognition under natural illumination could be especially valuable with respect to a scenario where an end-user makes a photograph of a material she/he wants a new piece of clothing to consist of (e.g. with her/his mobile phone), sends it to a cloud system and receives all the clothing agreeing with the specified material. In addition, the generation of synthetic training data similar to our approach might facilitate further investigations. Furthermore, reliable material recognition/retrieval techniques might become more valuable in industry as they allow to find either a certain material or a similar material in the databases from suppliers, to provide quality control or even to guide production processes based on a material description resulting from the material design process.

Future research in the context of material recognition might also be spent on the establishment of material recognition frameworks based on different types of training data including standard materials from commonly used databases such as texture databases, procedural models, databases with analytical BRDF models, databases with data-driven models such as the MERL BRDF database [MPBM03], the CUReT BTF database [DvGNK96], our UBO2014 database, BSSRDF models, or photo collections such as the Flickr Material Database [SRA09].

Another important objective for future research can be identified in the development of suitable material metrics to efficiently assess similarity or dissimilarity of materials based on distinctive material characteristics. This might allow a more practical material retrieval and material recognition. While similar efforts have long reached maturity in color science regarding the comparison of colors, the massive increase in physical degrees of freedom imposes significant challenges for the generalization of color metrics to general material appearance. Furthermore, the complexity of material appearance might impede that a single material metric is sufficient. Instead, a portfolio of metrics might have to be considered where the individual metrics are specific for certain material classes. The availability of a perceptual metric would allow for a perceptually uniform movement in the material space and thus also for the creation of perceptually uniform interpolation sequences. Especially designers would benefit from such techniques as they are offered to select the materials involved in their applications in a more intuitive way which might provide a connection between the virtual material design and reality.

Furthermore, material classification might be applied in the editing and fabrication domain, e.g. in the design of imitations where a designer searches for materials that are close to a certain query material. In many cases, exactly the same material is not available and similar materials have to be used instead. Material interpolation as e.g. approached in [RSK13], where plausible interpolations have been achieved even for materials with complex feature topology, spatially varying reflectance behavior and a mesostructure resulting in strong parallaxes, might also benefit from knowledge about how similar the materials to be interpolated are. It might even be possible to automatically specify between which materials an interpolation might be meaningful.

10.2.3 Efficient Automatic Acquisition

In the scope of this thesis, we have demonstrated that an efficient acquisition pipeline can be realized by a selection of the required acquisition techniques. With our investigations, we provide an inspiration towards consolidating more efficient acquisition pipelines that are required in industry. While we mainly focus on the efficient acquisition of the 3D shape of objects as well as on the acquisition of the geometric surface profile of flat material samples, we only rather briefly discussed that an automatic selection of appropriate reflectance acquisition techniques can be carried out in a similar way by an initial material recognition. In this thesis, we focused on rather flat material samples in this context. Certainly, further effort has to be spent to realize the suitable selection of acquisition techniques for objects with rather arbitrary 3D shapes. Not only an automatic segmentation has to be carried out as discussed in Section 10.2.2 to handle objects made of different materials, but also more different techniques for reflectance acquisition will have to be integrated into the acquisition pipeline, i.e. more different BRDF and SVBRDF models will have to be considered as well as more complex reflectance functions such as the BSSRDF. In addition, the material database including the respective annotations used for the selection of the appropriate acquisition techniques will have to be significantly extended due to the huge variation in appearance encountered for different materials.

In addition to the selection of appropriate geometry acquisition techniques such as structured light approaches, shape-from-specularity approaches as well as techniques for translucent and transparent materials, and the selection of adequate reflectance acquisition techniques based on different models such as BRDFs, SVBRDFs or BTFs, a future avenue of research might consider the automatic selection of suitable exposure times used during the acquisition. Furthermore, much more efficient algorithms for the measurement of material BTFs can be devised if certain priors such as a class-specific database of sample materials are present. As shown in [dBSHK14], BTFs can be acquired based on sparse measurements for a wide range of materials if prior knowledge in the form of a BTF database is available. In particular, the UBO2014 BTF database is used in the corresponding investigations. As similar materials exhibit similar characteristics in surface structure and surface reflectance, an initial clustering of the materials in the database is performed. This is followed by the fitting of linear patch-based models to each of the clusters and the reconstruction from sparse measurements can then be performed by solving a linear system of equations using a per-cluster sampling strategy derived from the models.

Furthermore, the concept of an efficient acquisition framework as presented in Chapter 9 might also significantly facilitate the acquisition of cloth. As the resolution of the fine-grained structures such as fibers or their hairiness is still below the resolution of the typical geometry acquisition techniques and reflectance acquisition techniques, volumetric fiber-based statistical yarn models and cloth models have been developed as discussed in e.g. [Sch13], where a pipeline to reverse engineer cloth and estimate a parametrized cloth model from a single image is introduced. The automatic estimation of yarn paths, yarn widths, their variation and a weave pattern provides fundamental insights that are used to accurately model the appearance of the original cloth sample. Furthermore, these properties derived from the input image provide a physically plausible basis that is fully editable using a few intuitive parameters.

BIBLIOGRAPHY

[AB91] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. 1991. 39

[Adl00] M. Adlers. *Topics in Sparse Least Squares Problems*. PhD thesis, Linkoping University, Linkoping, Sweden, 2000. 74

[AEB$^+$12] O. Aubreton, G. Eren, Y. Bokhabrine, A. Bajard, and F. Truchetet. Estimation of surface normal vectors based on 3D scanning from heating approach. In *Proceedings of SPIE*, volume 8290, pages 82900X–1 – 82900X–6, 2012. 37

[ALFG12] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 262–269, 2012. 173

[ALY08] D. G. Aliaga, A. J. Law, and Y. H. Yeung. A virtual restoration stage for real-world objects. *ACM Transactions on Graphics*, 27(5):149:1–149:10, 2008. 28

[AMKB04] S. Agarwal, S. P. Mallick, D. Kriegman, and S. Belongie. On refractive optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 483–494, 2004. 36

[And11] B. L. Anderson. Visual perception of materials and surfaces. *Current Biology*, 21(24):R978–R983, 2011. 106

[AVBSZ07] Y. Adato, Y. Vasilyev, O. Ben-Shahar, and T. Zickler. Towards a theory of shape from specular flow. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1–8, 2007. 31, 80

[AX08] D. G. Aliaga and Y. Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3D modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. 45, 46, 51

[AX10] D. G. Aliaga and Y. Xu. A self-calibrating method for photogeometric acquisition of 3D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(4):747–754, 2010. 45, 51, 66

[BAG12] M. Beljan, J. Ackermann, and M. Goesele. Consensus multi-view photometric stereo. In *Pattern Recognition*, volume 7476 of *Lecture Notes in Computer Science*, pages 287–296, 2012. 28, 32

[BARA06] A. Ben-Artzi, R. Ramamoorthi, and M. Agrawala. Efficient shadows for sampled environment maps. *Journal of Graphics Tools*, 11(1):13–36, 2006. 143

[BB90] A. Blake and H. Bulthoff. Does the brain know the physics of specular reflection? *Nature*, 343(6254):165–168, 1990. 106

[BBS10] T. L. Berg, A. C. Berg, and J. Shih. Automatic attribute discovery and characterization from noisy web data. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1, pages 663–676, 2010. 110

[BCJS06] N. Birkbeck, D. Cobzas, M. Jagersand, and P. Sturm. Variational shape and reflectance estimation under changing light and viewpoints. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 536–549, 2006. 66

[BEN03] M. Ben-Ezra and S. K. Nayar. What does motion reveal about transparency? In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1025–1032, 2003. 36

[BETvG08] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008. 26, 174

[BG06] G. J. Burghouts and J.-M. Geusebroek. Color textons for texture recognition. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1099–1108, 2006. 113

[BG09] G. J. Burghouts and J.-M. Geusebroek. Material-specific adaptation of color invariant features. *Pattern Recognition Letters*, 30(3):306–313, 2009. 117, 123, 124, 174, 212

[BHB11] J. Balzer, S. Holer, and J. Beyerer. Multiview specular stereo reconstruction of large mirror surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2537–2544, 2011. 32, 33, 159

[BHW11] M. Brown, G. Hua, and S. Winder. Discriminative learning of local image descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(1):43–57, 2011. 110

[BJK07] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision (IJCV)*, 72(3):239–257, 2007. 28

[BK69] B. Berlin and P. Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley and Los Angeles, 1969. 104

[Bla04] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1):231–243, 2004. 27

[BM12] J. T. Barron and J. Malik. Shape, albedo, and illumination from a single image of an unknown object. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 334–341, 2012. 135, 139

[BM13] J. T. Barron and J. Malik. Intrinsic scene properties from a single rgb-d image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17–24, 2013. 135, 139

[BP81] J. Beck and S. Prazdny. Highlights and the perception of glossiness. *Perception and Psychophysics*, 30(4):407–410, 1981. 106

[Bre01] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. 113, 145

[BRL97] N. Bhushan, A. R. Rao, and G. L. Lohse. The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images. *Cognitive Science*, 21(2):219–246, 1997. 104

[BS03] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 591–596, 2003. 32, 33, 80, 87, 89, 159

[BSBM05] J. Berzhanskaya, G. Swaminathan, J. Beck, and E. Mingolla. Remote effects of highlights on gloss perception. *Perception*, 34(5):565–575, 2005. 106

[BUSB13] S. Bell, P. Upchurch, N. Snavely, and K. Bala. Opensurfaces: A richly annotated catalog of surface appearance. *ACM Transactions on Graphics*, 32(4):111:1–111:17, 2013. 114, 135, 137, 174

[BUSB14] S. Bell, P. Upchurch, N. Snavely, and K. Bala. Material recognition in the wild with the materials in context database. *CoRR*, abs/1412.0623, 2014. 110, 113, 114, 135, 137, 174

[BW10] J. Balzer and S. Werling. Principles of shape from specular reflection. *Measurement*, 43:1305–1317, 2010. 31, 86

[BZM06] A. Bosch, A. Zisserman, and X. Munoz. Scene classification via pLSA. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 517–530, 2006. 144

[CD04] O. G. Cula and K. J. Dana. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision (IJCV)*, 59(1):33–60, 2004. 113, 120, 122, 123

[CG08] M. Crosier and L. D. Griffin. Texture classification with a dictionary of basic image features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2008. 109, 174

[CG10] M. Crosier and L. D. Griffin. Using basic image features for texture classification. *International Journal of Computer Vision (IJCV)*, 88(3):447–460, 2010. 109, 174

[CGS06] T. Chen, M. Goesele, and H.-P. Seidel. Mesostructure from specularity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1825–1832, 2006. 32, 34, 80, 81, 86, 159, 172

[Che95] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(8):790–799, 1995. 80, 89

[CHFE10] B. Caputo, E. Hayman, M. Fritz, and J.-O. Eklundh. Classifying materials in the real world. *Image and Vision Computing*, 28(1):150–163, 2010. 113

[CHM05] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1597–1604, 2005. 113, 134, 137, 145, 174

[CJ82] E. N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18(4):309–328, 1982. 30, 32

[CJ08] H.-S. Chung and J. Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. 29

[CK11] D. Cremers and K. Kolev. Multiview stereo and silhouette consistency via convex functionals over convex domains. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(6):1161–1174, 2011. 26, 66

[CLFS07] T. Chen, H. Lensch, C. Fuchs, and H.-P. Seidel. Polarization and phase-shifting for 3D scanning of translucent objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 34

[CLL07] J. Y. Chang, K. M. Lee, and S. U. Lee. Multiview normal field integration using level set methods. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 26, 28, 29, 80, 81, 82, 94

[CMK$^+$14] M. Cimpoi, S. Maji, I Kokkinos, S. Mohamed, and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3606–3613, 2014. 104, 105, 110, 113, 114, 137, 138, 174

[CMPP02] M. J. Chantler, G. McGunnigle, A. Penirschke, and M. Petrou. Estimating lighting direction and classifying textures. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 72.1–72.10, 2002. 112

[CSL08] T. Chen, H.-P. Seidel, and H. Lensch. Modulated phase-shifting for 3D scanning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. 34

[CT10] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2567–2573, 2010. 114, 117, 120

[CT11] F. Calakli and G. Taubin. SSD: Smooth signed distance surface reconstruction. *Computer Graphics Forum*, 30(7):1993–2002, 2011. 74, 90

[Cur10] P. Curd. *Anaxagoras of Clazomenae: Fragments and Testimonia - A Text and Translation with Notes and Essays*. G - Reference, Information and Interdisciplinary Subjects Series. University of Toronto Press, 2010. 3

[Dai09] Z. Dai. A Markov random field approach for multi-view normal integration. Master-Thesis, University of Hong Kong, 2009. 26, 28, 81, 94

[dBSHK14] D. den Brok, C. Steinhausen, M. B. Hullin, and R. Klein. Patch-based sparse reconstruction of material btfs. *Journal of WSCG*, 22(2):83–90, 2014. 175, 177

[dBSHK15] D. den Brok, C. Steinhausen, M. B. Hullin, and R. Klein. Multiplexed acquisition of bidirectional texture functions for materials. *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)*, 9398(14), 2015. 175

[dBWK15] D. den Brok, M. Weinmann, and R. Klein. Linear models for material BTFs. In *Eurographics Workshop on Material Appearance Modeling (to appear in)*, 2015. 175

[DC05] O. Drbohlav and M. Chantler. Illumination-invariant texture classification using single training images. In *Texture 2005: Proceedings of the International Workshop on Texture Analysis and Synthesis*, pages 31–36, 2005. 112

[Deb98] P. Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 189–198, 1998. 136, 142

[Deb13] P. Debevec. Light probe image gallery and high-resolution light probe image gallery. http://www.pauldebevec.com/, November 2013. Accessed on 1st November 2013. 142, 149, 151

[DFY$^+$11] K. Doerschner, R. W. Fleming, O. Yilmaz, P. R. Schrater, B. Hartung, and D. Kersten. Visual motion and the perception of surface material. *Current Biology*, 21(23):2010–2016, 2011. 106

[DHT$^+$00] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 145–156, 2000. 39

[DJV$^+$14] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 647–655, 2014. 110

[DLG13] Y. Dong, S. Lin, and B. Guo. *Material appearance modeling: A data-coherent approach*. Springer Berlin Heidelberg, Germany, 2013. 41

[DMZP14] B. Dong, K. D. Moore, W. Zhang, and P. Peers. Scattering parameters and surface normals from homogeneous translucent materials using photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2299–2306, 2014. 35, 159

[DN98] K. J. Dana and S. K. Nayar. Histogram model for 3D textures. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 618–624, 1998. 112

[DNvGK97] K. J. Dana, S. K. Nayar, B. van Ginneken, and J. J. Koenderink. Reflectance and texture of real-world surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 151–157, 1997. 112, 123, 135, 212

[DP11] A. Delaunoy and E. Prados. Gradient flows for optimizing triangular mesh-based surfaces: Applications to 3D reconstruction problems dealing with visibility. *International Journal of Computer Vision (IJCV)*, pages 100–123, 2011. 66

[DPB10] A. Delaunoy, E. Prados, and P. N. Belhumeur. Towards full 3D Helmholtz stereovision algorithms. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, volume 1, pages 39–52, 2010. 30, 66, 67

[DT05] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005. 109, 174

[DvGNK96] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink. Reflectance and texture of real world surfaces. Technical report, Department of Computer Science, Columbia University, Columbia, USA, 1996. Columbia University Technical Report CUCS-048-96. 117, 134, 137, 140, 174, 176

[EAM$^+$09] G. Eren, O. Aubreton, F. Meriaudeau, L. A. Sanchez Secades, D. Fofi, A. Teoman Naskali, F. Truchetet, and A. Ercil. Scanning from heating: 3D shape estimation of transparent objects from local surface heating. *Optics Express*, 17(14):11457–11468, 2009. 37

[EG08] M. Enzweiler and D. M. Gavrila. A mixed generative-discriminative framework for pedestrian classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. 135, 138

[ES04] C. Hernández Esteban and F. Schmitt. Silhouette and stereo fusion for 3D object modeling. *Comput. Vis. Image Underst.*, 96:367–392, 2004. 26, 66

[EVC08] C. Hernandez Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 30(3):548–554, 2008. 28, 66

[FA91] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(9):891–906, 1991. 109, 174

[FB81] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981. 51

[FCM$^+$08] Y. Francken, T. Cuypers, T. Mertens, J. Gielis, and P. Bekaert. High quality mesostructure acquisition using specularities. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2008. 32, 80, 81, 83, 86, 159

[FCMB09] Y. Francken, T. Cuypers, T. Mertens, and P. Bekaert. Gloss and normal map acquisition of mesostructures using gray codes. *Advances in Visual Computing*, 5876/2009:788–798, 2009. 86

[FDA03] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5), 2003. Article 3. 106

[FG87] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. *ISPRS Intercommission Workshop*, 1987. 109

[FG14] S. Fuhrmann and M. Goesele. Floating scale surface reconstruction. *ACM Transactions on Graphics*, 33(4):46:1–46:11, 2014. 52

[FH09] J. Filip and M. Haindl. Bidirectional texture function modeling: A state of the art survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(11):1921–1940, 2009. 140

[FJT11] R. W. Fleming, F. Jäkel, and Maloney L. T. Visual perception of thick transparent materials. *Psychological Science*, 22(6):347–368, 2011. 107

[FKIS02] R. Furukawa, H. Kawasaki, K. Ikeuchi, and M. Sakauchi. Appearance based object modeling using texture database: Acquisition, compression and rendering. In *Proceedings of the Eurographics Workshop on Rendering*, pages 257–266, 2002. 24

[Fle14] R. W. Fleming. Visual perception of materials and their properties. *Vision Research*, 94:62–75, 2014. 3, 106, 107, 108

[FTA04] R. W. Fleming, A. Torralba, and E. H. Adelson. Specular reflections and the perception of shape. *Journal of Vision*, 4(9):798–820, 9 2004. 106

[FWG13] R.W. Fleming, C. Wiebel, and K. Gegenfurtner. Perceptual qualities and material classes. *Journal of Vision*, 13(8), 2013. Article 9. 105, 106

[FY03] K. Fukui and O. Yamaguchi. Face recognition using multiviewpoint patterns for robot vision. In *Proceedings of the International Symposium of Robotics Research (ISRR)*, pages 192–201, 2003. 120, 121

[FZ07] V. Ferrari and A. Zisserman. Learning visual attributes. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*, pages 433–440, 2007. 110

[Gab46] D. Gabor. Theory of communication. *Journal of Institution of Electrical Engineers*, 93(26):429–457, 1946. 109, 174

[GAVN11] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. Structured light 3D scanning in the presence of global illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 713–720, 2011. 34, 172, 173

[GCHS05] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 1, pages 341–348, 2005. 29

[GCP$^+$10] A. Ghosh, T. Chen, P. Peers, C. A. Wilson, and P. Debevec. Circularly polarized spherical illumination reflectometry. *ACM Transactions on Graphics*, 29(6):162:1–162:12, 2010. 34

[GG03] L. Goddyn and P. Gvozdjak. Binary gray codes with long bit runs. *The Electronic Journal of Combinatorics*, 10:1–10, 2003. R27. 34

[GGSC96] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 43–54, 1996. 24, 39

[GILM07] B. Goldlucke, I. Ihrke, C. Linz, and M. Magnor. Weighted minimal hypersurface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(7):1194–1208, 2007. 37

[GL12] J. Gu and C. Liu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral brdf. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 797–804, 2012. 114

[GLL$^+$04] M. Goesele, H. P. A. Lensch, J. Lang, C. Fuchs, and H.-P. Seidel. Disco: Acquisition of translucent objects. *ACM Transactions on Graphics*, 23(3):835–844, 2004. 34

[GYN13] M. Gupta, Q. Yin, and S. K. Nayar. Structured light in sunlight. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 545–552, 2013. 173

[HBR11] D. Hu, L. Bo, and X. Ren. Toward robust material recognition for everyday objects. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1–11, 2011. 113

[HCFE04] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh. On the significance of real-world conditions for material classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 253–266, 2004. 113, 134, 137, 174

[HCP91] A. C. Hurlbert, B. G. Cumming, and A. J. Parker. Recognition and perceptual use of specular reflections. *Investigative Ophthalmology and Visual Science Supplement*, 32(4), 1991. 106

[HF13] M. Haindl and J. Filip. *Visual Texture: Accurate Material Appearance Measurement, Representation and Modeling*. Advances in Computer Vision and Pattern Recognition. Springer-Verlag New York Incorporated, 2013. 140

[HFI$^+$08] M. B. Hullin, M. Fuchs, I. Ihrke, H.-P. Seidel, and H. P. A. Lensch. Fluorescent immersion range scanning. *ACM Transactions on Graphics*, 27(3):87:1–87:10, 2008. 37

[HK06] A. Hornung and L. Kobbelt. Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 503–510, 2006. 66

[HLM06] Y.-X. Ho, M. S. Landy, and L. T. Maloney. How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, 6(5), 2006. Article 8. 106

[HLZ10] M. Holroyd, J. Lawrence, and T. Zickler. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Transactions on Graphics*, 29(4):99:1–99:12, 2010. 67

[HMI10] T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1157–1164, 2010. 30, 32

[HMJI09] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1234–1241, 2009. 66

[Hor70] B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA, 1970. 28

[HS81] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. 35

[HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, pages 147–151, 1988. 26, 109

[HS05] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27:1254–1264, 2005. 29

[HSKK96] S. Hata, Y. Saitoh, S. Kumamura, and K. Kaida. Shape extraction of transparent object using genetic algorithm. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 4, pages 684–688, 1996. 36

[HZ04] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision (2nd edition)*. Cambridge University Press, Cambridge, UK, 2004. 51, 52

[Ike81] K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on*

Pattern Analysis and Machine Intelligence (PAMI), 3(6):661–669, 1981. 32

[IKL$^+$10] I. Ihrke, K. N. Kutulakos, H. P. A. Lensch, M. Magnor, and W. Heidrich. Transparent and specular object reconstruction. *Computer Graphics Forum*, 29(8):2400–2426, 2010. 23, 24, 31, 35, 37

[IM05] I. Ihrke and M. Magnor. Adaptive grid optical tomography. In *Vision, Video, and Graphics 2005*, pages 141–148, 2005. 36, 37

[IMMY14] C. Inoshita, Y. Mukaigawa, Y. Matsushita, and Y. Yagi. Surface normal deconvolution: Photometric stereo for optically thick translucent objects. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 2, pages 346–359, 2014. 159

[ISM84] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-d object recognition. *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 806–808, 1984. 46

[Jak10] W. Jakob. Mitsuba renderer, 2010. http://www.mitsuba-renderer.org. 147, 149

[JCYS04] H. Jin, D. Cremers, A. J. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–36 – I–42, 2004. 66

[JDSP10] H. Jegou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3304–3311, 2010. 119, 121, 144, 145

[JSD$^+$14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. B. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia MM*, pages 675–678, 2014. 110

[JSJ10] M. Jehle, C. Sommer, and B. Jähne. Learning of optimal illumination for material classification. In *Pattern Recognition*, volume 6376 of *Lecture Notes in Computer Science*, pages 563–572, 2010. 114

[Kaj85] J. T. Kajiya. Anisotropic reflection models. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 15–21, 1985. 22

[KBBN09] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 365–372, 2009. 105

[KBBN11] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(10):1962–1977, 2011. 105

[KBH06] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the Eurographics Symposium on Geometry Processing (SGP)*, pages 61–70, 2006. 51, 52, 56, 57, 75

[KKDH07] M. Kazhdan, A. Klein, K. Dalal, and H. Hoppe. Unconstrained isosurface extraction on arbitrary octrees. In *Proceedings of the Eurographics Symposium on Geometry Processing (SGP)*, pages 125–133, 2007. 75

[KKH$^+$11] K. Kolev, N. Kirchgeíner, S. Houben, A. Csiszár, W. Rubner, C. Palm, B. Eiben, R. Merkel, and D. Cremers. A variational approach to vesicle membrane reconstruction from fluorescence imaging. *Pattern Recognition*, 44(12):2944–2958, 2011. 66

[KMA11] J. Kim, P. J. Marlow, and B. L. Anderson. The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision*, 11(9), 2011. Article 4. 106

[KPC10] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 538–551, 2010. 66

[KPDVG05] T. P. Koninckx, P. Peers, P. Dutre, and L. Van Gool. Scene-adapted structured light. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 611–618, 2005. 28

[KS05] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3D shape by light-path triangulation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1448–1455, 2005. 36, 159

[KS08] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3D shape by light-path triangulation. *International Journal of Computer Vision (IJCV)*, 76(1):13–29, 2008. 31, 36, 159

[KTF11] B. Kaneva, A. Torralba, and W. T. Freeman. Evaluation of image features using a photorealistic virtual world. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 2282–2289, 2011. 139

[KTM+02] A. Kanitsar, T. Theussl, L. Mroz, M. Sramek, A. V. Bartroli, B. Csebfalvi, J. Hladuvka, D. Fleischmann, M. Knapp, R. Wegenkittl, P. Felkel, S. Rottger, S. Guthe, W. Purgathofer, and M. E. Groller. Christmas tree case study: Computed tomography as a tool for mastering complex real world objects with applications in computer graphics. In *Proceedings of the Conference on Visualization*, pages 489–492, 2002. 37

[LA09] M. I. A. Lourakis and A. A. Argyros. SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software*, 36(1):1–30, 2009. 51

[Lau94] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 16:150–162, February 1994. 24

[LBN08] A. Ladikos, S. Benhimane, and N. Navab. Multi-view reconstruction using narrow-band graph-cuts and surface normal optimization. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 15.1–15.10, 2008. 66

[Lev44] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics*, II(2):164–168, 1944. 72

[LF12] W. Li and M. Fritz. Recognizing materials from virtual examples. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 4, pages 345–358, 2012. 113, 135, 139, 144, 149

[LF14] W. Li and M. Fritz. Learning multi-scale representations for material classification. *[arXiv]*, 2014. 110, 113, 114, 174

[LFCK12] L. Liu, P. Fieguth, D. Clausi, and G. Kuang. Sorted random projections for robust rotation-invariant texture classification. *Pattern Recognition*, 45(6):2405–2418, 2012. 109, 174

[LG14] C. Liu and J. Gu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral BRDF. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(1):86–98, 2014. 114

[LH96] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 31–42, 1996. 39

[Lin98] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision (IJCV)*, 30(2):79–116, 1998. 109

[LK81] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, volume 2, pages 674–679, 1981. 35

[Llo57] S. P. Lloyd. Least squares quantization in PCM. Technical report, Bell Laboratories, 1957. 33, 119, 121, 144, 145

[Llo82] S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28:129–137, 1982. 33, 119, 121, 144, 145

[LM99a] T. Leung and J. Malik. Recognizing surfaces using three-dimensional textons. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1010–1017, 1999. 109, 113, 116, 119, 174

[LM99b] C.-J. Lin and J. J. Moré. Incomplete Cholesky factorizations with limited memory. *SIAM Journal on Scientific Computing*, 21(1):24–45, 1999. 74

[LM01] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision (IJCV)*, 43(1):29–44, 2001. 109, 113, 116, 119, 174

[Low99] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1150–1157, 1999. 174

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60:91–110, 2004. 26, 109, 174

[LPC$^+$00] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 131–144, 2000. 34

[LSAR10] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 239–246, 2010. 109, 113, 114, 118, 122, 137, 144, 146

[LSP06] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2169–2178, 2006. 109

[LTBEB10] Z. Lu, Y.-W. Tai, M. Ben-Ezra, and M. S. Brown. A framework for ultra high resolution 3D imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1205–1212, 2010. 66

[LWDC10] M. Liu, K.-Y. K. Wong, Z. Dai, and Z. Chen. Specular surface recovery from reflections of a planar pattern undergoing an unknown pure translation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, volume 2, pages 137–147, 2010. 32

[LYG13] C. Liu, G. Yang, and J. Gu. Learning discriminative illumination and filters for raw material classification with optimal projections of bidirectional texture functions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1430–1437, 2013. 113, 114, 134, 137, 174

[Mac67] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967. 33, 119, 121, 144, 145

[MBK05] G. Müller, G. H. Bendels, and R. Klein. Rapid synchronous acquisition of geometry and BTF for cultural heritage artefacts. In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST)*, pages 13–20, 2005. 24, 52

[MC09] A. Maki and R. Cipolla. Obtaining the shape of a moving object with a specular surface. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 39.1–39.10, 2009. 30, 32

[MCUP02] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of* *the British Machine Vision Conference (BMVC)*, pages 36.1–36.10, 2002. 109

[MHP+07] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eurographics Symposium on Rendering (EGSR)*, pages 183–194, 2007. 34

[MK05] N. J. W. Morris and K. N. Kutulakos. Dynamic refraction stereo. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1573–1580, 2005. 35, 159

[MK07] N. J. W. Morris and K. N. Kutulakos. Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1–8, 2007. 36, 159

[MKA12] P. J. Marlow, J. Kim, and B. L. Anderson. The perception and misperception of specular surface reflectance. *Current Biology*, 22(20):1909–1913, 2012. 106

[MKSN12] C. Mertz, S. J. Koppal, S. Sia, and S. Narasimhan. A low-power structured light sensor for outdoor scene reconstruction and dominant material identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 15–22, 2012. 173

[MMS+04] G. Müller, J. Meseth, M. Sattler, R. Sarlette, and R. Klein. Acquisition, synthesis and rendering of bidirectional texture functions. In *Eurographics 2004 State of the Art Reports*, pages 69–94, 2004. 21, 22, 41

[Mor80] H. P. Moravec. *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Stanford University, Stanford, CA, USA, 1980. 109

[Mot10] I. Motoyoshi. Highlight-shading relationship as a cue for the perception of translucent and transparent materials. *Journal of Vision*, 10(9), 2010. Article 6. 107

[MPBM03] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, 2003. 176

[MRA+12] F. Meriaudeau, R. Rantoson, K. Adal, D. Fofi, and C. Stolz. Nonconventional imaging systems for 3D digitization of transparent objects: Shape from polarization in the IR and shape from visible fluorescence induced UV. In *International Topical Meeting on Optical Sensing and Artificial Vision (OSAV)*, pages 34–40, 2012. 37

[MS04] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision (IJCV)*, 60(1):63–86, 2004. 109

[MS05] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(10):1615–1630, 2005. 109

[MS13] R. Munoz-Salinas. ArUco: Augmented reality library from the university of cordoba. http://www.uco.es/investiga/grupos/ava/node/26, 2013. Accessed on 1st February 2013. 109

[MSSE^{+}10] F. Meriaudeau, L. A. Sanchez Secades, G. Eren, A. Ercil, F. Truchetet, O. Aubreton, and D. Fofi. 3-d scanning of nonopaque objects by means of imaging emitted structured infrared patterns. *IEEE Transactions on Instrumentation and Measurement*, 59(11):2898–2906, 2010. 37

[MT09] K. Mikolajczyk and T. Tuytelaars. Local image features. *Encyclopedia of Biometrics*, pages 939–943, 2009. 109

[Mur90] H. Murase. Surface shape reconstruction of an undulating transparent object. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 313–317, 1990. 35, 36

[Mur92] H. Murase. Surface shape reconstruction of a nonrigid transport object using refraction and motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(10):1045–1052, 1992. 35, 36

[NKGR06] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics*, 25:935–944, 2006. 34, 57, 173

[NRDR05] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics*, 24:536–543, 2005. 66

[NRH^{+}77] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometrical considerations and nomenclature for reflectance. National Bureau of Standards Monograph #160, U.S. Department of Commerce, 1977. 39

[NS98] S. Nishida and M. Shinya. Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A*, 15(12):2951–2965, 1998. 106

[NWR08] D. Nehab, T. Weyrich, and S. Rusinkiewicz. Dense 3D reconstruction from specularity consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008. 32, 33, 36, 80, 83, 89, 159

[NWSS90] S. K. Nayar, L. E. Weiss, D. A. Simon, and A. C. Sanderson. Specular surface inspection using structured highlight and Gaussian images. *IEEE Transactions on Robotics and Automation*, 6(2):208–218, 1990. 32

[OBUvdS13] C. Osendorfer, J. Bayer, S. Urban, and P. van der Smagt. Unsupervised feature learning for low-level local image descriptors. *arXiv preprint arXiv:1301.2840*, 2013. 110

[OJR03] M. Osadchy, D. W. Jacobs, and R. Ramamoorthi. Using specularities for recognition. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1512–1519, 2003. 112

[ON12] G. Oxholm and K. Nishino. Shape and reflectance from natural illumination. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 528–541, 2012. 135, 139

[OPM02] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(7):971–987, 2002. 109, 174

[Pak12] A. Pak. Recovering shapes of specular objects in motion via normal vector map consistency. In *Proceedings of SPIE*, volume 8493, pages 84930T–1 – 84930T–8, 2012. 32

[Pak13] A. Pak. Reconstruction of specular surfaces via probabilistic voxel carving. In *Proceedings of SPIE*, volume 8791, pages 87911B–1 – 87911B–8, 2013. 33

[PD07] F. Perronnin and C. R. Dance. Fisher kernels on visual vocabularies for image categorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 144

[PDG$^+$08] S. Padilla, O. Drbohlav, P. R. Green, A. Spence, and M. J. Chantler. Perceived roughness of 1/f$^\beta$ noise surfaces. *Vision Research*, 48(17):1791–1797, 2008. 106

[PJW$^+$11] L. Pishchulin, A. Jain, C. Wojek, M. Andriluka, T. Thormählen, and B. Schiele. Learning people detection models from few training samples. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1473–1480, 2011. 135, 138

[PK02] S. C. Pont and J. J. Koenderink. Bidirectional texture contrast function. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 4 of *Lecture Notes in Computer Science*, pages 808–822, 2002. 106

[RB06] S. Roth and M. J. Black. Specular flow and the recovery of surface structure. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1869–1876, 2006. 31, 80

[RK09] R. Ruiters and R. Klein. Heightfield and spatially varying brdf reconstruction for materials with interreflections. *Computer Graphics Forum (Proceedings of Eurographics)*, 28(2):513–522, 2009. 30

[RSC87] W. T. Reeves, D. H. Salesin, and R. L. Cook. Rendering antialiased shadows with depth maps. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, volume 21, pages 283–291, 1987. 143

[RSFM10a] R. Rantoson, C. Stolz, D. Fofi, and F. Meriaudeau. 3D reconstruction of transparent objects exploiting surface fluorescence caused by UV irradiation. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 2965–2968, 2010. 37

[RSFM10b] R. Rantoson, C. Stolz, D. Fofi, and F. Meriaudeau. Non contact 3D measurement scheme for transparent objects using UV structured light. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 1646–1649, 2010. 37

[RSGS10] M. Ritz, M. Scholz, M. Goesele, and A. Stork. High resolution acquisition of detailed surfaces with lens-shifted structured light. In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST)*, pages 1–8, 2010. 28, 44, 64

[RSK13] R. Ruiters, C. Schwartz, and R. Klein. Example-based interpolation and synthesis of bidirectional texture functions. *Computer Graphics Forum (Proceedings of Eurographics)*, 32(2):361–370, 2013. 142, 176

[Rui13] R. A. Ruiters. *Data-Driven Analysis and Interpolation of Optical Material Properties*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, 2013. 161

[SCD$^+$06] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 519–528, 2006. 26, 93

[Sch01] C. Schmid. Constructing models for content-based image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 39–45, 2001. 109, 174

[Sch13] K. M. N. Schroeder. *Visual Prototyping of Cloth*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, 2013. 178

[Sch14] J. C. Schwartz. *Acquisition, Transmission and Rendering of Objects with Optically Complicated Material Appearance*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, 2014. 163, 166, 167, 168, 213

[SCP05] S. Savarese, M. Chen, and P. Perona. Local shape from mirror reflections. *International Journal of Computer Vision (IJCV)*, 64(1):31–67, 2005. 32

[SFC$^+$11] J. Shotton, A. W. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1297–1304, 2011. 135, 139

[SFS09] N. Salamati, C. Fredembach, and S. Süsstrunk. Material classification using color and NIR images. In *Proceedings of the Color and Imaging Conference (CIC)*, pages 216–222, 2009. 174

[SGS10] M. Stark, M. Goesele, and B. Schiele. Back to the future: Learning shape models from 3D CAD data. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 106.1–106.11, 2010. 135, 139

[SL00] D. Skocaj and A. Leonardis. Range image acquisition of objects with non-uniform albedo using structured light range sensor. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 778–781, 2000. 27, 30

[SLRA13] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson. Recognizing materials using perceptually inspired features. *International Journal of Computer Vision (IJCV)*, 103(3):348–371, 2013. 109, 113, 122, 137, 144, 145, 146

[SN13] G. Schwartz and K. Nishino. Visual material traits: Recognizing per-pixel material context. In *Proceedings of the Color and Photometry in Computer Vision Workshop (Workshop held in conjunction with ICCV 2013)*, pages 883–890, 2013. 105, 110, 111, 113, 114, 174

[SP05] S. N. Sinha and M. Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 1, pages 349–356, 2005. 66

[SPB04] J. Salvi, J. Pagès, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37:827–849, 2004. 27

[SRA09] L. Sharan, R. Rosenholtz, and E. H. Adelson. Material perception: What can you see in a brief glance? *Journal of Vision*, 8, 2009. 103, 114, 134, 137, 174, 176

[SS02] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision (IJCV)*, 47:7–42, 2002. 26

[SS03] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 195–202, 2003. 27, 30, 46

[SSS08] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from Internet photo collections. *International Journal of Computer Vision (IJCV)*, 80(2):189–210, 2008. 51

[SSW$^+$14] C. Schwartz, R. Sarlette, M. Weinmann, M. Rump, and R. Klein. Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the University of Bonn. *Sensors*, 14(5):7753–7819, 2014. 57, 83, 117, 126, 140, 161, 162, 163, 175

[SSWK13] C. Schwartz, R. Sarlette, M. Weinmann, and R. Klein. DOME II: A parallelized BTF acquisition system. In *Proceedings of the Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition*, pages 25–31, 2013. 57, 162, 163

[STD09] G. Sansoni, M. Trebeschi, and F. Docchio. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009. 27

[Ste57] H. Steinhaus. Sur la division des corps matériels en parties. *Bulletin de l'Académie Polonaise des Sciences, Classe III*, 4:801–804, 1957. 33, 119, 121, 144, 145

[Str83] G. Strang. Maximal flow through a domain. *Mathematical Programming*, 26(2):123–143, 1983. 66

[Str10] G. Strang. Maximum flows and minimum cuts in the plane. *Journal of Global Optimization*, 47(3):527–535, 2010. 66

[SVTA10] A. C. Sankaranarayanan, A. Veeraraghavan, O. Tuzel, and A. K. Agrawal. Specular surface reconstruction from sparse reflection correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1245–1252, 2010. 31, 80

[SVZ12] K. Simonyan, A. Vedaldi, and A. Zisserman. Descriptor learning using convex optimisation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 7572 of *Lecture Notes in Computer Science*, pages 243–256, 2012. 110

[SVZ14] K. Simonyan, A. Vedaldi, and A. Zisserman. Learning local feature descriptors using convex optimisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(8):1573–1585, 2014. 110

[SWN88] A. C. Sanderson, L. E. Weiss, and S. K. Nayar. Structured highlight inspection of specular surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 10(1):44–55, 1988. 32, 80

[SWRK11] C. Schwartz, M. Weinmann, R. Ruiters, and R. Klein. Integrated high-quality acquisition of geometry and appearance for cultural heritage. In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST)*, pages 25–32, 2011. 43, 57, 67, 162, 163, 175

[TBH06] B. Trifonov, D. Bradley, and W. Heidrich. Tomographic reconstruction of transparent objects. In *Eurographics Symposium on Rendering (EGSR)*, pages 51–60, 2006. 37

[TFCRS11] W. Thompson, R. Fleming, S. Creem-Regehr, and J. K. Stefanucci. *Visual Perception from a Computer Graphics Perspective*. A. K. Peters, Ltd., Natick, MA, USA, 2011. 106

[TFG$^+$13] B. Tunwattanapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Transactions on Graphics*, 32(4):109:1–109:12, 2013. 33

[TGZ08] A. T. Targhi, J.-M. Geusebroek, and A. Zisserman. Texture classification with minimal training images. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008. 135, 138

[TLGS05] M. Tarini, H. P. A. Lensch, M. Goesele, and H.-P. Seidel. 3D acquisition of mirroring objects using striped patterns. *Graphical Models*, 67(4):233–259, 2005. 33, 80

[TM08] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008. 108, 109

[TMHF00] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, pages 298–372, 2000. 50

[TNM04] J. T. Todd, J. Farley Norman, and E. Mingolla. Lightness constancy in the presence of specular highlights. *Psychological Science*, 15(1):33–39, 2004. 106

[Tro95] M. Trobina. Error model of a coded-light range sensor. Technical report, Communication Technology Laboratory, ETH Zentrum, Zürich, 1995. 46, 57, 84

[TSDSR] The Stanford 3D Scanning Repository. `http://graphics.stanford.edu/data/3Dscanrep/`. 75, 91, 94, 212

[TVG12] R. Timofte and L. Van Gool. A training-free classification framework for textures, writers, and materials. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1–12, 2012. 146

[Vap95] V. N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995. 145

[VBvdWV11] E. Vazquez, R. Baldrich, J. van de Weijer, and M. Vanrell. Describing reflectances for color segmentation robust to shadows, highlights, and textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 33(5):917–930, 2011. 175

[vdSGS10] K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(9):1582–1596, 2010. 109

[vGKD99] B. van Ginneken, J. J. Koenderink, and K. J. Dana. Texture histograms as a function of irradiation and viewing direction. *International Journal of Computer Vision (IJCV)*, 31(2-3):169–184, 1999. 112

[VHTC07] G. Vogiatzis, C. Hernández, P. H. S. Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(12):2241–2246, 2007. 66

[Vis] Visual Geometry Group (University of Oxford). Texture classification. Accessed on 1st February 2014. 111, 123

[VZ02] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 255–271, 2002. 113

[VZ03] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 691–698, 2003. 113

[VZ04] M. Varma and A. Zisserman. Unifying statistical texture classification frameworks. *Image and Vision Computing*, 22(14):1175–1183, 2004. 113

[VZ09] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(11):2032–2047, 2009. 109, 113, 145

[War92] G. J. Ward. Measuring and modeling anisotropic reflection. *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 26(2):265–272, 1992. 29

[WD04] J. Wang and K. J. Dana. Hybrid textons: Modeling surfaces with reflectance and geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 372–378, 2004. 139

[WdBKK15] M. Weinmann, D. den Brok, S. Krumpen, and R. Klein. Appearance capture and modeling. In *SIGGRAPH Asia 2015 Courses*, pages 4:1–4:1, 2015. 19, 140

[Wei13] M. Weinmann. Visual features - from early concepts to modern computer vision. In *Advanced Topics in Computer Vision*, Advances in Computer Vision and Pattern Recognition, pages 1–34. Springer London, 2013. 108, 109

[WFM08] G. Wendt, F. Faul, and R. Mausfeld. Highlight disparity contributes to the authenticity and strength of perceived glossiness. *Journal of Vision*, 8, 2008. Article 14. 106

[WGK14] M. Weinmann, J. Gall, and R. Klein. Material classification based on training data synthesized using a BTF database. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 3, pages 156–171, 2014. 113, 133

[WGS$^+$11] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 703–717, 2011. 28

[WGSD09] O. Wang, P. Gunawardane, S. Scher, and J. Davis. Material classification using BRDF slices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2805–2811, 2009. 112

[WHON97] T.-T. Wong, P.-A. Heng, S.-H. Or, and W.-Y. Ng. Image-based rendering with controllable illumination. In *Proceedings of the Eurographics Workshop on Rendering*, pages 13–22, 1997. 135

[WI93] Z. Wang and S. Inokuchi. Determining shape of specular surfaces. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA)*, pages 1187–1194, 1993. 33

[WK15a] M. Weinmann and R. Klein. Advances in geometry and reflectance acquisition. In *SIGGRAPH Asia 2015 Courses*, pages 1:1–1:71, 2015. 19, 41, 140

[WK15b] M. Weinmann and R. Klein. Material recognition for efficient acquisition of geometry and reflectance. In *Computer Vision - ECCV 2014 Workshops*, volume 8927 of *Lecture Notes in Computer Science*, pages 321–333, 2015. 115, 155

[WLDW11] C. Wu, Y. Liu, Q. Dai, and B. Wilburn. Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(8):1082–1095, 2011. 28, 66

[WLGK16] M. Weinmann, F. Langguth, M. Goesele, and R. Klein. Advances in geometry and reflectance acquisition. In *Computer Graphics Forum (Proceedings of Eurographics)*, 2016. (accepted). 19, 41, 140

[WLZ$^+$09] H. Wang, M. Liao, Q. Zhang, R. Yang, and G. Turk. Physically guided liquid surface modeling from videos. *ACM Transactions on Graphics*, 28(3):90:1–90:11, 2009. 37

[Woo80] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 28, 90

[WORK13] M. Weinmann, A. Osep, R. Ruiters, and R. Klein. Multi-view normal field integration for 3D reconstruction of mirroring objects. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 2504–2511, 2013. 33, 79

[WRO$^+$12] M. Weinmann, R. Ruiters, A. Osep, C. Schwartz, and R. Klein. Fusing structured light consistency and Helmholtz normals for 3D reconstruction. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1–12, 2012. 31, 63, 77

[WSRK11] M. Weinmann, C. Schwartz, R. Ruiters, and R. Klein. A multi-camera, multi-projector super-resolution framework for structured light. In *Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 397–404, 2011. 28, 30, 43, 53, 75, 77

[WWRJ16] M. Weinmann, M. Weinmann, F. Rottensteiner, and B. Jutzi. Acquisition and automatic characterization of scenes - from point clouds to features and objects. In *ISPRS Congress 2016 Tutorials*, 2016. (accepted). 19

[XA09] Y. Xu and D. G. Aliaga. An adaptive correspondence algorithm for modeling scenes with strong interreflections. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 15(3):465–480, 2009. 57

[XHE$^+$10] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3485–3492, 2010. 109, 119

[YAC06] T. Yu, N. Ahuja, and W. C. Chen. SDG cut: 3D reconstruction of non-Lambertian objects using graph cuts on surface distance grid. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2269–2276, 2006. 66

[YBT10] J. Yuan, E. Bae, and X.-C. Tai. A study on continuous max-flow and min-cut approaches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2217–2224, 2010. 66, 73, 82, 90

[YFM98] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 318–323, 1998. 120, 121

[YIX07] M. Yamazaki, S. Iwata, and G. Xu. Dense 3D reconstruction of specular and transparent objects using stereo cameras and phase-shift method. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 570–579, 2007. 32, 159, 172

[YWT$^+$11] S.-K. Yeung, T.-P. Wu, C.-K. Tang, T. F. Chan, and S. Osher. Adequate reconstruction of transparent objects on a shoestring budget. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2513–2520, 2011. 33, 159, 172

[YY11] Y. Yoshiyasu and N. Yamazaki. Topology-adaptive multi-view photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1001–1008, 2011. 66

[YYT$^+$13] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T.F. Chan. Outdoor photometric stereo. In *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, 2013. 173

[Zai11] Q. Zaidi. Visual inferences of material changes: color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6):686–700, 2011. 106

[ZBK02] T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision (IJCV)*, 49(2-3):215–227, 2002. 30, 67, 70, 72

[ZGB89] A. Zisserman, P. Giblin, and A. Blake. The information available to a moving observer from specularities. *Image and Vision Computing*, 7(1):38–42, 1989. 32

[Zha00] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(11), 2000. 109

[ZHK$^+$03] T. E. Zickler, J. Ho, D. J. Kriegman, J. Ponce, and P. N. Belhumeur. Binocular Helmholtz stereopsis. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1411–1417, 2003. 30

[ZM00] J. Y. Zheng and A. Murata. Acquiring a complete 3D model from specular motion under the illumination of circular-shaped light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(8):913–920, 2000. 32

[ZMLC10] Z. Zheng, L. Ma, Z. Li, and Z. Chen. An extended photometric stereo algorithm for recovering specular object shape and its reflectance properties. *Computer Science and Information Systems*, 7(1):1–12, 2010. 29

[ZMLS07] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision (IJCV)*, 73(2):213–238, 2007. 144

DATA SOURCES

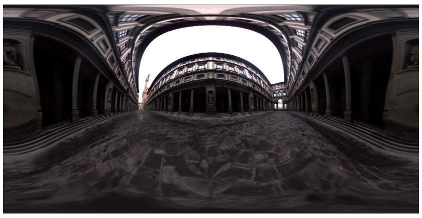
The following freely available data used in this thesis have been taken from external sources.

Stanford Bunny
The Stanford 3D Scanning Repository
http://graphics.stanford.edu/data/3Dscanrep/

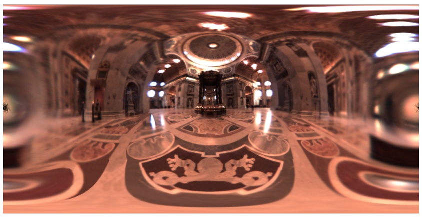
Lucy
The Stanford 3D Scanning Repository
http://graphics.stanford.edu/data/3Dscanrep/

Happy Buddha
The Stanford 3D Scanning Repository
http://graphics.stanford.edu/data/3Dscanrep/

Kitchen at 2213 Vine St Light Probe
©1999 Paul Debevec
http://www.debevec.org/Probes/



The Uffizi Gallery, Florence, Light Probe
©1999 Paul Debevec
http://www.debevec.org/Probes/



St. Peter's Basilica, Rome, Light Probe
©1999 Paul Debevec
http://www.debevec.org/Probes/

Dining room of the Ennis-Brown House, Los Angeles, California, Light Probe
©2008-2013 USC Institute for Creative Technologies
http://gl.ict.usc.edu/Data/HighResProbes/

LIST OF FIGURES

List of Tables