UNIVERSITY OF VERONA
DEPARTMENT OF COMPUTER SCIENCE

SURFACE ANALYSIS AND VISUALIZATION FROM MULTI-LIGHT IMAGE COLLECTIONS

Tinsae Gebrechristos Dulecha
Supervisor: Prof. Andrea Giachetti

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science of the University of Verona.

May 14, 2021

This thesis is dedicated to my family

for their endless love, support and encouragement

Abstract

Multi-Light Image Collections (MLICs) are stacks of photos of a scene acquired with a fixed viewpoint and a varying surface illumination that provides large amounts of visual and geometric information. Over the last decades, a wide variety of methods have been devised to extract information from MLICs and have shown its use in different application domains to support daily activities. In this thesis, we present methods that leverage a MLICs for surface analysis and visualization.

First, we provide background information: acquisition setup, light calibration and application areas where MLICs have been successfully used for the research of daily analysis work. Following, we discuss the use of MLIC for surface visualization and analysis and available tools used to support the analysis. Here, we discuss methods that strive to support the direct exploration of the captured MLIC, methods that generate relightable models from MLIC, non-photorealistic visualization methods that rely on MLIC, methods that estimate normal map from MLIC and we point out visualization tools used to do MLIC analysis.

In chapter 3 we propose novel benchmark datasets (RealRTI, SynthRTI and SynthPS) that can be used to evaluate algorithms that rely on MLIC and discusses available benchmark for validation of photometric algorithms that can be also used to validate other MLIC-based algorithms. In chapter 4, we evaluate the performance of different photometric stereo algorithms using SynthPS for cultural heritage applications. RealRTI and SynthRTI have been used to evaluate the performance of (Neural)RTI method.

Then, in chapter 5, we present a neural network-based RTI method, aka NeuralRTI, a framework for pixel-based encoding and relighting of RTI data. In this method using a simple autoencoder architecture, we show that it is possible to obtain a highly compressed representation that better preserves the original information and provides increased quality of virtual images relighted from novel directions, particularly in the case of challenging glossy materials.

Finally, in chapter 6, we present a method for the detection of crack on the surface of paintings from multi-light image acquisitions and that can be used as well on single images and conclude our presentation.

Acknowledgments

I first thank the almighty God for everything. I then must express my deepest gratitude to my supervisor professor Andrea Giachetti, for his scientific guidance, encouragement and fruitful discussion during my Ph.D. journey.

I also would love to thank my co-authors: Ruggero Pintus; Irina Ciortan and Enrico Giobetti for the fruitful discussion we had during my Ph.D. journey. And, special thanks go to my wife Giovanna Bertacche, son Leo Tinsae Dulecha and my friends: Leul, Seyum and Teddy; please accept my sincere gratitude for being a good friend; moreover, for the unforgettable moments, we had in Campo and Baum (coffee area).

I also would like to thank the Department of Computer Science of the University of Verona and the Ministry of Education, University and Research of Italy for financing my Ph.D. study.

I thank my external reviewers: Prof. Holly Rushmeier and Prof. Massimiliano Corsini for the time they spent carefully reading the thesis and for their insightful comments and suggestions.

Last but not least, I thank my family for the unconditional love and support, especially my mother (Etagegn bishaw) and my father (Gebrechristos Dulecha).

Preface

This dissertation is submitted in fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Verona. The thesis presents methods that rely on Multi-Light Image Collections (MLICs) for surface analysis and visualization. The first chapter introduces background knowledge: acquisition, light calibration and applications of MLICs. The second chapter presents the use of MLICs for surface visualization and analysis [106]; it has been appeared in Computer Graphics Forum (CGF). The third chapter presents a novel real (RealRTI) and synthetic (SynthRTI and SynthPS) MLIC benchmark datasets. The RealRTI is acquired using light domes and a free-form acquisition setup. Whereas, SynthRTI and SynthPS are generated with Physically-Based rendering (PBR) techniques. SynthRTI and RealRTI have been used to validate the performance of the NeuralRTI method [38]. The fourth chapter evaluates the performance of selected photometric stereo methods using SynthPS dataset for cultural heritage applications that can be used to derive suggestions or even guidelines for the practical reconstruction of normal maps and surfaces from MLIC data; it has been appeared in Eurographics Workshop on Graphics and Cultural Heritage 2020 (EG GCH 2020). The fifth chapter discusses a novel neural network-based Reflectance Transformation Imaging (NeuralRTI) method that is used to compress and relight MLICs (RTI) dataset. This work [38] has been presented at Computer Graphics International (CGI) 2020 conference and published in The Visual Computer (TVC) journal. Finally, the last chapter presents the Convolutional Neural Network-based method for the detection of crack from the surface of painting [37]. This work has appeared in a EG GCH 2019.

Introduction

A set of images contains more information about the scene than a single image. If the image set is structured by systematic changes to illumination direction, the characteristics of objects in the scene may be inferred. This set of images referred as Multi-Light Image Collections (MLICs) [43] or Reflectance Transformation Imaging (RTI) Stacks [87]. MLICs are an effective means to gather detailed information on the shape and appearance of objects. The basic idea of the approach is to visually characterize objects by capturing multiple images of the surface of interest from a fixed point of view, changing the illumination direction at each shot. Such a collection of samples, typically arranged in image stacks, provides massive amounts of visual data that can be analyzed to extract information and knowledge on the shape and appearance of a surface.

The tremendous improvement of quality and price of controllable lighting and digital photography has made the acquisition of high-resolution MLICs practical and affordable using many different physical setups. As a result, a wide variety of applications are using MLICs as a major means to non-destructively gather information on scenes and objects at many scales, as well as to provide users with useful visualization tools for object analysis. These applications benefit from an array of computational tools that have been devised in different domains, such as feature detection and enhancement [43], reconstruction of normal and 3D shapes [2, 154], extraction of surface appearance and material behavior [35, 151], illustrative visualization, [53, 80], BRDF fitting [93] and creation of relightable images [38, 81].

In this thesis, we present methods that exploit MLICs to perform surface analysis and visualization task, propose a method which creates a relightable image from MLICS, using neural network, and method for the detection of crack from the surface of painting based on MLICs and convolutional neural networks. We start by providing structured background information of MLIC. Following, we discuss how MLIC can be used for surface visualization and analysis, and what are the available visualizations tools to perform analysis. Then, we present a novel MLIC benchmark dataset, neural network-based Re-

flectance Transformation Imaging (RTI) method and a method that detects crack from the surface of paintings exploiting MLIC.

The first chapter provides background information on Multi-light image collections. Here, we discuss different acquisition setup, categorizing them into two main categories. Then, we discuss different light calibration techniques. Some of the techniques for visualizing and analyzing MLICs can indeed work without any other information than the captured images. However, most of the visualization and analysis methods assume that some information is available on camera and lights. We conclude Chapter 1 by discussing application areas where MLICs have been successfully used for the research of daily analysis work.

According to a recent survey [106] MLIC can be used for surface analysis and visualization and plays the main role in supporting daily activities in different application domains. These roles are usually supported using different MLIC visualization tools. In chapter 2, we discuss the use of MLIC for surface visualization and analysis and available tools used to support the analysis. To be specific, we discuss methods that strive to support the direct exploration of the captured MLIC, methods that generate relightable models from MLIC, non-photorealistic visualization methods that rely on MLIC, normal maps and visualization tools used to do MLIC analysis.

One of the challenges in the Multi-Light Image Collections (MLIC) community, is the lack of benchmarks for evaluation. Usually, MLIC based methods evaluated using specific homemade datasets or a few publicly available benchmark (for example DiLiGent, [132]). The main problem of these types of a benchmark is that they are not representative of those typically captured in the real-world applications (e.g. quasi-planar surfaces made of heterogeneous materials with a wide range of metallic and specular behaviors and non-uniform materials) of MLIC acquisition. Thus, the availability of specific benchmarks for different surface analysis tasks will be fundamental to validate methods and develop guidelines for the use of MLIC visualization software in different domains. To this end, in chapter 3 we propose benchmarks that can be used to evaluate algorithms that rely on MLIC. And in chapter 4, we evaluate the performance of different photometric stereo algorithms using SynthPS for cultural heritage applications.

Reflectance Transformation Imaging [40, 81, 87] is a popular computational photography technique, allowing the capture of rich representations of surfaces including geometric details and local reflective behavior of materials. This method is widely used in the cultural heritage and material science domains to characterize relieved surfaces. In chapter 5, we propose a neural network-based RTI method, aka NeuralRTI, a framework for pixel-based encoding and relighting of RTI data. In this method using a simple autoencoder architecture, we show that it is possible to obtain a highly compressed representation that better preserves the original information and provides increased quality of virtual images relighted from novel directions, particularly in the case of challenging glossy materials.

Cracks represent an imminent danger for painted surfaces that needs to be alerted before degenerating into more severe aging effects, such as color loss. Automatic detection

of cracks from painted surfaces' images would be therefore extremely useful for art conservators; however, classical image processing solutions are not effective to detect them, distinguish them from other lines or surface characteristics. A possible solution to improve the quality of crack detection exploits Multi-Light Image Collections (MLICs). Hence, in chapter 6 we present a method for the detection of crack on the surface of paintings from multi-light image acquisitions and that can be used as well on single images. The proposed method is based on a processing pipeline able to automatically identify crack regions in single and multi-light images. The pipeline is based on a (multi-light) edge detection and a Convolutional Neural Network-based labeling of image patches around edges.

To summarize, the main contributions of this thesis are:

- It provides structured MLIC background information: acquisition, light calibration, applications and visualization tools.
- It provides a structured presentation of MLIC for surface visualization and analysis. It presents methods that support the direct exploration of the captured MLIC, methods that generate a fused single-image from MLIC, methods that generate a relightable model from MLIC and non-photorealistic visualization techniques that rely on MLIC.
- It proposes a novel real (RealRTI) and synthetic (SynthRTI and SynthPS) MLIC benchmark datasets that can be used for validation of MLIC-related algorithms (for instance: photometric stereo, Reflectance Transformation Imaging (RTI), image relighting, material classification).
- It evaluates the performance of different photometric stereo algorithms using a novel dataset (SynthPS) for cultural heritage applications.
- It proposes a neural network-based Reflectance Transformation Imaging (RTI) method, aka NeuralRTI, a framework for pixel-based encoding and relighting of RTI data and evaluate its performance using RealRTI and SynthRTI benchmarks.
- Finally, it presents a pipeline for the detection of crack on the surface of paintings from multi-light image acquisitions and that can be used as well on single images. The method is based on single or multi-light edge detection and on a custom Convolutional Neural Network able to classify image patches around edge points as crack or non-crack, trained on MLIC data.

Publications

I.M. Ciortan, **Dulecha, T.G.**, A. Giachetti, R. Pintus, A. Jaspe-Villanueva, and E. Gobbetti. Artworks in the spotlight: characterization with a multispectral led dome. In *IOP Conference Series: Materials Science and Engineering*, volume 364, page 012025. IOP Publishing, 2018.

Ruggero Pintus, **Dulecha, Tinsae**, Alberto Jaspe, Andrea Giachetti, Irina Mihaela Ciortan, and Enrico Gobbetti. Objective and subjective evaluation of virtual relighting from reflectance transformation imaging data. In *GCH*, pages 87–96, 2018.

Ruggero Pintus, **Dulecha, Tinsae Gebrechristos**, I Ciortan, Enrico Gobbetti, and Andrea Giachetti. State-of-the-art in multi-light image collections for surface visualization and analysis. In *Computer Graphics Forum*, volume 38, pages 909–934. Wiley Online Library, 2019.

Dulecha, Tinsae, Andrea Giachetti, Ruggero Pintus, Irina-Mihaela Ciortan, Alberto Jaspe Villanueva, and Enrico Gobbetti. Crack detection in single-and multi-light images of painted surfaces using convolutional neural networks. In *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association, 2019.

Dulecha, Tinsae G, Filippo A Fanni, Federico Ponchio, Fabio Pellacini, and Andrea Giachetti. Neural reflectance transformation imaging. *The Visual Computer*, pages 1–14, 2020.

Dulecha, Tinsae G., Ruggero Pintus, Enrico Gobbetti, and Andrea Giachetti. Synthps: a benchmark for evaluation of photometric stereo algorithms for cultural heritage applications. In *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association, 2020.

Dulecha, Tinsae Gebrechristos, Andrea Dall'Alba, and Andrea Giachetti. Mlic-synthetizer: a synthetic multi-light image collection generator. In *STAG*, pages 105–106, 2019.

Contents

List of Figures

List of Tables

CHAPTER 1

Background

Multi-Light Image Collections (MLICs) are sets of photos of a scene acquired from the same fixed viewpoint but with varying lighting conditions. These sets of images contain more information about the scene than one image alone. The motivation for acquiring these sets of images comes from the assumption that by structuring the sets of images changing an illumination direction, it is possible to infer the characteristics of objects in the scene and to visually characterize them. MLICs are commonly acquired in several practical applications for surface analysis and visualization. We present an overview of them in Section:1.3), but, first, we give in Section 1.1 some technical background on their acquisition and calibration. To capture multiple images of a surface requires a specific setup procedure integrated with general-purpose equipment such as a digital camera, light source, based on the individual objective and requirement. Usually, once the acquisition process is finished, the relative position of the light for the acquired object surface is estimated and saved for further use.

1.1 Acquisitions setups

So far, several multi-light acquisition setups, with different cost requirements, dimensions and running environments geared around material appearance measurement (photometric stereo, relighting, shape reconstruction, BRDF measurement) have been introduced. These setups differ for several features, but, generally speaking, they can be classified into a free-from and light dome. The first is based on a single light source moved around the object,

typically manually and the second uses multiple light sources placed into a fixed structure (e.g. light dome).

1.1.1 Free-form acquisition

The free-form acquisition, as we can guess from the name, it is the form of acquisition where the light source moves freely, not fixed to a specific position. In this type of acquisition, usually, we have one light source but to acquire MLIC, the light source moves around the scene in such a way that the object is lit by light from different directions. Several free-form acquisition setups have been introduced so far, with different cost requirements, and objectives. As the light source position is not constrained, this type of setup allows an arbitrary and potentially dense direction sampling, but methods to estimate light directions in the camera space are mandatory for subsequent image processing. Special targets inserted in the scene are thus necessary and are the features characterizing the different approaches proposed together with the light source employed (see figure1.5). In the past, various, this type of acquisitions with a little difference have been presented by different researchers.

Masselus et al. [83] presented an acquisition setup, to captures the reflectance field of an object using a free-moving, hand-held light source. This setup contains a digital camera, fixed during acquisition, a hand-held light source, and four diffuse white spheres for light calibration (see Figure:1.1). Winnemöller et al [153] proposed an acquisition setup similar to [83] with minor modifications to record fixed viewpoint reflectance fields by waving a light source around the scene. Similarly, Chen et al [19] proposed an acquisition setup targeted to estimate mesostructure for shiny samples from observations of specular reflections from a hand waved light source. In this setup, Jenoptik ProgRes C14 digital camera, a 5Watt Luxeon Star white LED as point light source, a checkerboard, and four specular spheres to estimate the light source position are used.

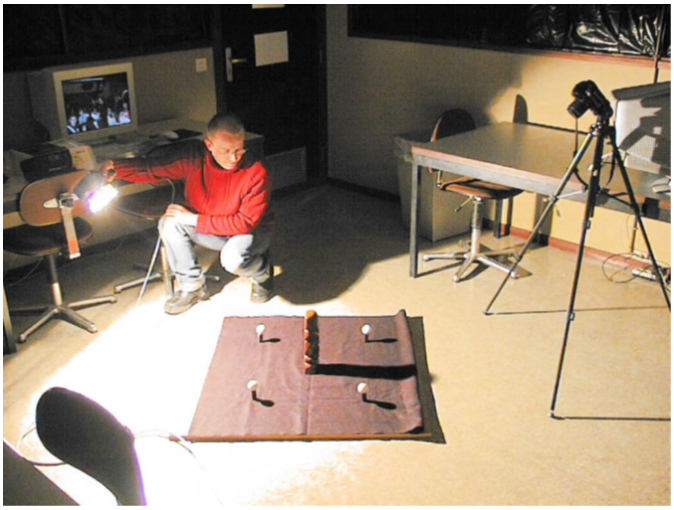


Figure 1.1: *An handheld light acquisition setup by Masselus, et al. [83]*

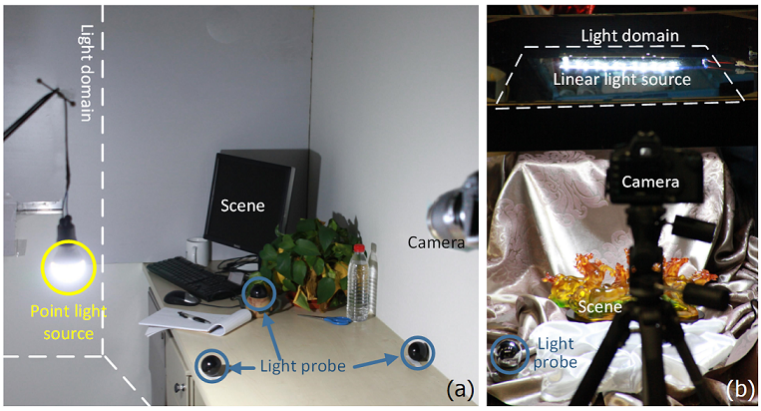


Figure 1.2: *An acquisition setup by Ren et al. [123]*

Giachetti et al [48] presented a free-form hand-held setup to obtain shape and appearance information of captured surfaces, as well as to derive compact relightable representations of them. The acquisition setup consists of a stationary DSLR digital camera, a hand-held light source, some calibration targets (a Lambertian white background/frame and four black reflective spheres), and the object being acquired positioned on a planar support perpendicular to the camera axis. Goldman et al. [51] proposed a setup for microscopic acquisition to investigate archaeological textiles by utilizing micro-RTI technique.

Ren et al.[123] introduced an acquisition setup to acquire appearance information of a surface using a deep learning approach. In this setup reflective spheres are placed near the target scene to recover the light source position. A point light source and linear light source composed of 12 LEDs are used to light the scenes. Besides, to constrain light movement to a 2D, the glass panel is placed above the scene to keep the light source in contact with the glass (see Figure:1.2).

Several issues related to the planning of handheld light MLIC captures have been addressed by Dellepiane et al. [30]. Here the authors extended the acquisition setup used to generate relightable images (PTM) from photographs to large objects which in turn solve issues related to the size of the object and conditions of acquisitions. They studied the degradation of the quality of the PTMs concerning the number and position of lights used to acquire the PTM. To simplify the manual light placements, they also developed a software tool, called PTM Planner which let define the properties of the "virtual" light dome to visually check its correctness and to automatically generate the coordinates for the light placements.

1.1.2 Light domes

In contrast to the free-from setup, the most effective and widely used technique is a dome, acrylic hemisphere, based acquisition technique. Unlike the hand-held acquisition setup, in the dome-based acquisition, both the light and camera are placed in a fixed position (camera mounting point and the lamp positions are fixed). Thus, leads to higher stability and precision. Due to its nature of the setup, Dome based acquisition has many advantages,

including speed of acquisition, accuracy, and repeatability. But it has also limitations, in particular, the maximum object size, cost, portability, and difficulty of adapting to the site topography.

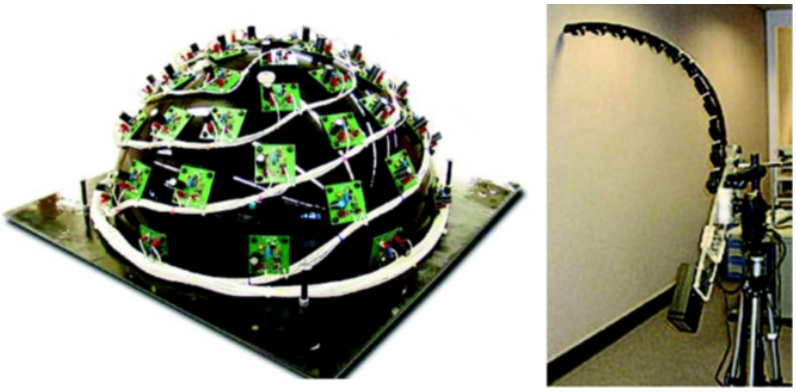
Several dome solutions have been introduced so far, with different cost requirements, dimensions and running environments [20, 24, 52, 97]. Malzbender et al. [81] built a hemispherical device with 50 light sources (see Figure 1.4(a)), suitable for sampling small objects (nearly 15 cm.). It is a 90 cm diameter black plastic hemisphere, with fifty evenly distributed strobe lights mounted such that they illuminate the hemispheric dome's interior. The digital camera is positioned at the top of the hemisphere and photographs the subject through a viewport cut in the dome. The same authors propose also a second setup (Figure 1.4(b)), designed for larger objects. Here, a 90 degrees arc 1.50m in diameter is mounted with 12 strobe lights facing towards the center of the arc. One end of the arc is connected to a circular bearing race in the shape of a torus. This allows the arc to spin in a 360 degrees circle around the bearing race.

Palma et al. [97] introduced a flexible dome solution composed by four aluminum shells that can easily be assembled and disassembled to simplify the transport. It has 116 cold white LEDs and an overhead high-resolution reflex camera. The dome is computer-controlled to allow a completely automatic acquisition by synchronizing the switching on/off each LED with the shutter of the camera. Pitard et al. [111] designed a multi-view dome solution for the case of quality inspection of small objects. In this solution, they focused on the upper dome of the spherical structure using only a single camera for gathering and processing reflectance transformation imaging.

Hameeuw [52] proposed a portable dome solution extremely easy to use. It is almost a one-click solution that can provide a straightforward RTI acquisition pipeline. The dome is equipped with a 5 million pixel camera or 29 million monochromatic sensors; it mounts 260 white LED light sources around an 80cm diameter dome. The main design purpose is to make it easy to assemble/disassemble the dome in less than half an hour. This requirement makes it operable both in museum collections and other in-situ scenarios. Recently, other domes have been presented that use both visible and invisible light wavelengths. One example is the micro-dome with multispectral RTI capability presented by the RICH team [72], [147]. It is equipped with 228 different LED light sources. Those LEDs are divided into five different spectra: ultraviolet (365 nm), blue (460 nm), red (523 nm), green (623 nm), and infrared (850 nm). A 28 megapixels monochromatic sensor is mounted on top of the dome.

Recently, Ciortan et al. [24] introduced a novel dome solution using low-cost hardware and reasonable portability making it suitable for both lab and on-site acquisitions. It is equipped with 156 multi-spectral LEDs evenly distributed across 52 light positions over a 60cm diameter hemisphere. The light sources cover 5 bands: two narrows in the ultraviolet (centered at 395nm) and infrared (centered at 850nm) regions and one broad in the visible to collect the RGB signals of the Nikon D810 DSLR, a camera with the infrared cut-off filter removed. Besides, it also includes a user-friendly interface to control light and camera. The acquisition setup of Ciortan et al. is shown in figure 1.3.

Figure 1.3: *A multispectral light-dome based acquisition setup by Ciortan et al. [24]*



(a) **(b)**

Figure 1.4: *Malzabender etal acquistion setup [81]. a) PTM acquistion dome. b) PTM acquistion arc.(photos: hp labs taken from [30])*

1.2 Light Calibration

The relative positions and orientations of the individual components to the object surface, as well as device-specific characteristics, are important for the accurate acquisition of geometry and reflectance [151]. As a consequence, there is a need for geometric calibration, using certain known calibration objects or possibly also objects with the unknown shape(see Figure:1.5), that allows obtaining a geometric relationship between the involved illuminants, imaging sensors and the object surface during the acquisition.

The calibration of real illumination is a well-known topic in Computer Vision, and, specifically, in the Photometric Stereo (PS) field [2]. While some methods try to implicitly consider real light sources within the particular PS framework [102, 160], others are more focused on the explicit calibration of various lighting properties. Some methods make assumptions on the light form factor, e.g., near point light [3] or linear light source [144]

and try to exploit the illuminated scene to extract the light position and direction. For instance, Ahmad et al. [3] exploit diffused maxima regions in the framed model and derive from them the light directions. Others perform calibration by sampling light on calibration targets of known shape and albedo (e.g., reflective spheres or diffuse targets). Corsini et al. [27] use high-dynamic-range images of two reflective balls to acquire the spatially-varying illumination of a real-world scene, and it focuses more on the environment light effect rather than on the computation of a per-pixel light direction and intensity. Ackermann et al. [1] present a study and validation through error statistics of both a forward and backward geometric point light source calibration by using sets of different numbers of reflective spheres. Although it proposes a very simple and robust way to compute light direction, it considers a point light model without considering non-uniform light intensity.

Other methods strongly rely on a specific, fixed light form factor (e.g., LED light [104, 159], and model the intensity with the corresponding fall-off due to both distance and angle to the light principal axis. Xie et al. [159] also consider vignetting effects. Unfortunately, those methods do not apply to the case of a general variable illumination due to non-ideal lamps or lenses. Giachetti et al. [48] presents a novel light calibration technique. In their approach, they use multiple spheres to estimate a light direction field and use measures on a planar white target to estimate the intensity of each light ray, infilling missing data with a low-degree interpolation, thus reconstructing an approximation of the entire light field illuminating the scene. Some works thus try to cope with non-uniform intensity without imposing an analytical light model [5, 140]: they use a flat reference object with known albedo to sample an arbitrary lighting vector field and to calibrate it using a flat-fielding approach. They don't use polynomial interpolation, but they exploit measured spatially varying intensities to compensate for the input images and to convert the problem into a standard collimated case. Differently, from Giachetti et al. [48], they require different acquisitions for the calibration step and the actual capture; this is possible only with a fixed light configuration, but it does not apply to a more general free-form, hand-held multi-light acquisition.

McDonald [79] employed three different calibration methods. The first method employed for determining the lamp coordinates was to take images of the shadow cast onto graph paper by a vertical steel pin when illuminated by each light(figure:1.5 first image). The two-dimensional coordinates of the tip of the shadow can be determined simply by inspection of the image, and trigonometry applied. The second method, which they believe gives a more precise determination of the lamp coordinates, close-range photogrammetry was employed, with the aid of the Vision Measurement System (VMS). To apply VMS to measure the lamp coordinates in the dome, retro-reflective targets were stuck onto four well-defined positions around each of the 64 flash lamps, and a photogrammetric project was undertaken. The third one is the method that realizes using the flash themselves as a recognizable target (Figure:1.5 2^{nd} and 3^{rd} images).

1.3 Applications

MLIC techniques have been mixed and matched to solve several problems in a variety of application domains, and they have been also integrated into user-friendly software tools (section 2.6) aimed at supporting both visualization and analysis.

The visual analysis of MLICs is at the basis of several applications and tasks, ranging

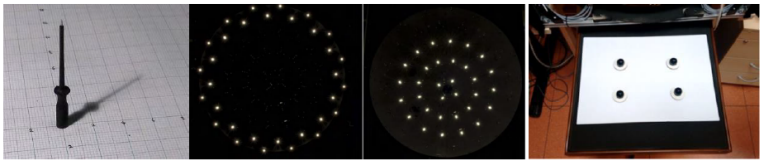


Figure 1.5: *Example of calibration objects. The first three calibration objects used by McDonald [79] and last one the calibration object used by Giachetti et al. [48]*

from pure visualization to geometric analysis [108]. In the sequel, we provide a structured overview of the areas where MLICs have been successfully used for daily visualization and analysis work. Table 1.1 presents a schematic view of the distribution of the techniques presented in chapter 2 to the main application domains they are employed in.

1.3.1 Cultural Heritage

Cultural Heritage (CH) appears, by far, to be the application domain where MLICs are more popular. This is due to both the versatility of the capturing approach (capable of working at a large number of scales) and the large quantity of information that can be gathered about object surfaces by using relatively low-cost equipment. Moreover, using moving lights fits well with traditional CH analysis means, and it is thus well received by CH experts. Nowadays MLIC based acquisition is considered a well-assessed digitization method in CH, and specific surveys and tutorials have been dedicated to them [88, 148].

Since their introduction, relightable images based on MLICs (section 2.3) have been used for the analysis of CH objects like tablets, statues [36, 45, 81, 87, 89, 129], paintings [95, 116], coins [97, 116], and manuscripts [150]. The interactive relighting is particularly appreciated by CH experts as it also enables the simulation of the typical raking light surface inspection [110]. Another common practice in the CH community is related to the use of illustrative enhancements (Sec.2.4), such as *Diffuse gain*, *Specular enhancement*, and *Unsharp masking*, to intensify the detail perception on tablets [53, 98], petroglyphs or written stones [91, 145], mural graffiti [28], coins [67], or, generally, metallic alloys [78]. In this context, a major research need is the enhancement of legibility and readability of written texts or engraved symbols [40], and the processing of MLICs can lead to the extraction of meaningful lines and inscriptions on tablets [11, 100]. Further, an advanced application to automatically recover line drawings of inscriptions has been demonstrated on a collection of Mesopotamian clay cones [52]. The combination of the surface penetration given by the infrared light together with the non-photorealistic exaggeration facilitated by the multi-directional light images has proven effective for recovering invisible texts of the Dead Sea Scrolls [15], where the deterioration of the parchment caused its blending with the black iron-gall ink. Furthermore, by using RBF interpolation (section 2.1), the raking light analysis is enabled, and certain engravings around the margins of the lamina gain more visibility.

Another application of MLICs coveted by the CH end-users is the visualization of the dynamic interplay between light and the complex surfaces of artworks characterized by heterogeneous optical behaviors. In a recent work [44], the authors manage to trace back the original visual impression of mosaic tesserae and early prints by unmixing the glossy

details from the diffuse component as a result of fitting the light transport function formulated as a polynomial (section 2.3). In this manner, they obtain a layered representation of the artwork's reflectance. Moreover, to allow the end-users to customize the digital surrogate according to their artistic view, the gloss layer is enhanced in a non-photorealistic fashion (section 2.4) and generated synthetically according to a Phong reflection model. Through the user input, parameters such as the degree of shininess, diffusiveness, and specularities can be interactively modified to render different degrees of glossiness. The separation of the specular component has been proposed as well by Macdonald [78] as a way to improve the quality of CH object relighting. The importance of the specular reconstruction for the visualization of relighted paintings' surface is demonstrated also by Thanikachalam et al. [142], where the method based on the reconstruction of light transport matrix with a compressive sensing approach (section 2.3) preserves the visual richness of paintings and allows for the visualization of how the specular behavior migrates over the painting's surface. Pitard et al. [112] show that a better interactive relighting model could not only improve the quality of the specular component on a jewel visualization but also allows for a better understanding of texture relief on a wall painting.

The analysis of the rich data in MLICs has given promising results in revealing the manufacturing techniques of a craftsman or the painting style of an artist. This application is especially relevant for reclaiming information of unknown or unstudied heritage, as well as very old or lost-and-found historical objects. Newman [92], based on the specular enhancement visualizations (section 2.4), distinguishes between three types of negative profiles (incisions and striations) in the creation and finishing of archaeological bone artifacts from El Zotz assemblage. At the same time, Artal and Klausmeyer [6] propose a method that uses non-photorealistic enhancements and feature maps in the case of Greek red-figure vases (section 2.4) to guide the end-user in distinguishing between various types of decoration lines with positive profiles, as well as their order of application at the fabrication moment. Similarly, others spot different orientation patterns in the clay molding process of Execration Egyptian statuettes made of unbaked clay [146]. Analogously, in painted items, it is possible to render the relief and orientations of the brushstrokes related to a style or an artist. In addition, many methods aim at detecting particular pigments through direct local MLIC interpolation in the light space (section 2.1), by mixing the multispectral data into a single false-color visualization [147, 148] (section 2.2), or by directly analyzing histograms of the per-pixel reflection maps (section 2.1). In this context, MLICs have also been used to recover lost pigments and to revive faded colorings by using non-photorealistic renderings [146]. Finally, fibers in historical textiles with different materials and patterns of decay can be visually detected by using magnified multi-light images under UV radiation [51] (section 2.2).

MLICs can be maneuvered as a measurement tool in assessing the conservation condition of CH materials. For instance, single image visualization and non-photorealistic enhancements based on MLICs of coins [67] indicate the need for material cleaning and show whether a conservation treatment was applied successfully or not. Similarly, the incorporation of profile measurements on an exaggerated shaded view of manuscripts has been shown to provide useful visual annotation concerning the analyzed documents and their medium [147]. In addition to single image visualizations and dramatic enhancements, by deriving feature maps that target crack formation, Manfredi et al. [82] propose a quantitative approach for monitoring damages in paintings. With a resembling underly-

ing procedure, MacDonald [79] measured the extent of cracks in a fresco at the National Gallery of London.

1.3.2 Natural Sciences

MLICs are a well-recognized imaging technique for natural sciences as well [88]. In biology, MLIC-based relightable image models (section. 2.3) are used to support animal species identification [55], while MLIC-based image fusion (section. 2.2) is used for detail visualization of leaves or flowers [43, 121, 167]. Non-photorealistic visualization methods combined with feature extraction have also been successfully used for the visual analysis of biological surfaces [149] and leaf venations [165]. In paleontology, MLIC visualization methods have mainly two functionalities: detail visualization and whitening. The low visibility of images of fossils is hampering visual analysis. Mostly, this factor comes due to low color contrast and a corroded relief. To solve this problem, the classic RTI framework is often used to illustrate subtle features from fossils, assisting in discovering hidden details [54]. MLIC data are used not only to create interactively relightable images of the fossils [80, 81] (section 2.3), but also to simulate fossil whitening, a task performed to get information regarding their morphology. Normally, this task is done manually using various techniques such as an airbrush or chemical agent (ammonium chloride). Hammer and Spocova [54] show that this can be done digitally using MLIC-based photorealistic visualization, with competitive performance to traditional more invasive techniques. Enhancements like edge detection have been used as well to increase the perception of fossil structures [100].

1.3.3 Industry

In industrial manufacturing, the inspection of an object's surface is often performed to control whether the final product meets predefined tolerances. In recent years, major efforts have been made to replace, wherever possible, manual or physical testing with a virtual visual inspection. MLIC visualization is emerging as a promising solution since it allows for the reliable acquisition of materials, their high-quality visualization, and an exchange of that information for industrial monitoring and design [157]. Enhanced single image visualization based on MLICs can be used to better understand the structure of mechanical parts [121]. Pitard et al. [113, 114] have used a MLIC photorealistic visualization technique for the detection and the analysis of visual anomalies on challenging metallic surfaces. This technique is derived from DMD [111] (section 2.3), and automates the visual inspection by introducing a rotation-invariant representation of the reflectance characteristics of the inspected surface. Smith and Stamp [135] also use MLIC based photorealistic visualization technique for quality inspection of textured ceramic tiles, while Rump et al. [126] employed MLICs visualization for metallic car paint inspection and design. In the context of quality control and investigation of electronic reliability, MLICs captured with the Scanning Electron Microscope (SEM) [99] are used to produce meaningful visualization of micro-and nano-details on the surface of electronic circuits; Pintus et al. [109] demonstrate an application to the visualization of damaged solid-state electron devices.

1.3.4 Underwater inspection

An image-based surface analysis in the underwater environment is a difficult task, due to factors such as water clarity/turbidity, excessive algae, and other participating media. This highly affects the light phenomena, and usually results in dark, low contrast, and bluish tone-image. A specialized approach of MLICs applied to that type of environment has been proposed by Selmo et al. [130]. Authors use RTI to study underwater CH objects, e.g. details from historic shipwrecks. They also perform a study on the effects of turbidity on relighting quality. Another issue in underwater MLIC processing is related to light refraction effects. Fan et al. [42] propose a flat refractive model to cope with this phenomenon by computing the mapping of virtual and real points; they can remove distortions in underwater acquisitions both for 3D reconstruction and MLIC visualization.

1.3.5 Medical imaging

The application of MLICs in the area of medical science strives to increase the readability of medical images. It is mostly used in the type of examination where the internal organ such as the esophagus, stomach, and upper part of the small intestine is visually inspected, often in real-time during examinations [25, 103]. This is typically achieved by modifying existing instruments (e.g., endoscopes [39]) to perform multi-light acquisitions, and by fusing several images to help in visualization. The typical approach is to reconstruct normal and remove highlights in the visualization [103]. By using a simple MLIC capture with few light sources and single image enhanced visualization it is possible to assist doctors in the understanding of the 3D structure of anatomical parts in endoscopic views [121]. MLIC analysis is also used for the 3D reconstruction of different organs. For instance, Lv et al. [76] used Photometric Stereo to retrieve the shape of a tongue, which is used for an accurate diagnosis. To support visualization, the reconstructed organ is mapped with the surface albedo. MLIC-based visualization is also used in dermatology, which often uses interactive visual inspection to support diagnostic tasks. In this area, a photorealistic visualization with fitted relightable models (section 2.3) has been proposed for pigment lesion analysis [139], or to find biomarkers for early skin cancer detection [138]. Sun et al. [139] have shown that MLIC-based non-photorealistic enhancement (section 2.4) can generate an efficiently encoded and reasonably complete representation of skin appearance, which assists in the analysis and diagnosis of pigmented skin lesions.

1.4 Summary

To sum up, in this chapter we have discussed the different contexts of MLIC. We started by discussing how MLIC can be acquired, what are the different acquisition setup protocols utilized by the MLIC community, categorizing them into a free-form and light dome. Then, we have discussed different existing, MLIC-based, light calibration techniques (which is fundamental information for most MLIC based algorithms). Finally, we have discussed in detail the application areas where MLIC is used to support daily activities. The areas include cultural heritage, natural science, industry, underwater inspection and medical imaging.

	Cultural Heritage	Natural Sciences	Industry	Underwater	Medical Imaging
Direct exploration (Sec. 2.1)	Interpretation/Monitoring [47]; Dissemination/Inspection [47, 116, 129].				
Single-image data fusion (Sec. 2.2)	Enhanced perception [24, 28]; Showing the invisible [24, 146, 147, 150]; Pigments [28, 146, 147, 150].	Enhanced perception [43, 121, 167]; Automatic illustration [43, 121, 167]; Dissemination/Inspection [43].	Enhanced perception [121].		Endoscopy [121]; Shape and reflectance analysis [138, 139].
Relightable images (Sec. 2.3)	Surface defects [79]; Material characterization [78]; Interpretation/Monitoring [51, 88, 92]; Dissemination/Inspection [36, 44, 51, 79, 81, 88, 97, 112, 142].	Enhanced perception [81, 88]; Species identification/Interpretation [55]; Dissemination/Inspection [81, 88].	Surface Quality Inspection [111, 112, 135].		Shape and reflectance analysis [138, 139].
Non-photorealistic enhancements (Sec. 2.4)	Enhanced perception [15, 28, 40, 53, 81, 87, 91, 95, 98, 145]; Showing the invisible [15, 91]; Pigments [28, 47]; Surface defects [95]; Interpretation/Monitoring [6, 15, 40, 67, 92, 146, 147]; Dissemination/Inspection [6, 40, 67, 81, 87, 89, 95, 98, 145, 147]	Enhanced perception [54, 80, 81]; Automatic illustration [54]; Species identification/Interpretation [55]; Dissemination/Inspection [54, 80, 81].		Enhanced perception [130].	

Table 1.1: ***Techniques vs Major Application Domains.*** *For each class of MLIC analysis and visualization methods, the major use cases for each application domain are presented.*

CHAPTER 2

Multi-Light Image Collections for Surface Visualization and Analysis

Multi-Light Image Collections (MLICs) contain plenty of information that can provide large amounts of visual and geometric data if extracted properly. Over the past decades, a wide variety of methods have been devised to extract information from MLICs. In this chapter, we focus on those techniques from the point of view of the data and interaction means that they make available to users for analysis and visualization.

In Section one, we present methods that strives to support the *direct exploration* of the captured MLIC (Section 2.1). These techniques typically exploit a light-space parameterization to offer a user-interface to browse the input dataset, eventually performing interpolation to provide continuity of motion, or to simulate in-between lighting configurations. Little processing or enhancements are applied in these techniques, which are mostly used as components of larger data exploration systems. Minimal calibration is required here, most of the time, limited to an approximation of the spatial configuration of lights. In section two, we present group of techniques that generates a *fused single-image* (Section 2.2) that maximizes for each output pixel the amount of information present in the input MLIC. This is generally done by an optimization process, and the resulting image is static, and not used in interactive settings. User interaction is typically limited to changing parameters to guide the detail enhancement process. Moreover, many of them can be applied without the knowledge of the illumination pattern, thus limiting calibration needs.

In section three, we present techniques used to generate *relightable models*. These

techniques typically rely on, at least approximately, calibrated data, and compute per-pixel representations that can be exploited at visualization time to generate novel illumination patterns. A method in this class, possibly the most popular one, constructs fitted representations that approximate the light transport function, or perform a light matrix domain fitting (Section 2.3). A user interface is generally provided to control the simulation of novel lighting, which is used in exploration applications (e.g., using simulated raking lights to enhance particular features). In contrast to the *direct exploration* methods, the lighting used during the interactive task is not limited to be an interpolation of the ones used for the acquisition.

In section four, we present non-photorealistic visualization techniques. These techniques help in gaining insight on objects through visual effects (Section 2.4). These methods are rarely applied for enhancing specific aspects of the raw images (e.g., color discontinuities and color frequencies at different levels of resolution), while, most often, they are combined with relightable image methods to support an improved visual interpretation of the scene of interest in an interactive setting.

In section five, we present different photometric stereo algorithms. A method focused on estimating the surface normal given MLICS. Finally, we present available interactive MLIC-based visualization tools used to perform MLIC analysis.

2.1 Direct exploration of acquired data

Direct exploration approaches aim at providing a framework to browse original data, by performing little processing except, eventually, for the computation of in-between images to simulate a continuous sampling of the light space. One of the main challenges is the need for dealing with large amounts of data at exploration time.

In the case of original image navigation, no processing is needed on the image stack, so the only issue is how to offer interactive access to the displayed images, which is done either by a simple discrete selector or by finding the nearest neighbor in a continuous interface for specifying the light position [79].

2.1.1 (PCA-compressed) Radial Basis Function

In the direct interpolation scenario, Radial Basis Functions (RBF) [14] interpolation has been typically used to produce a relighted image with a new virtual light from the MLIC stack [47]. Given the N images corresponding to the closest neighbors of the new light direction, RBF interpolation is achieved by computing the parameters that define a sum of N radial functions. The reflectance I is expressed as:

$$I\left(\vec{l}\right) = \sum_{i=1}^{N} \alpha_i e^{\frac{\|\vec{l}-\vec{l}_i\|^2}{R^2}} \tag{2.1}$$

where α_i is estimated per pixel by solving a least-squares problem.

A recent objective and subjective evaluation have shown, in particular, how RBF interpolation is capable of better visualizing the behavior of complex materials for in-between views with respect to classic fitting approaches [105] (Fig. 2.1). However, the need to access large amounts of data makes the method very computation and memory-intensive.

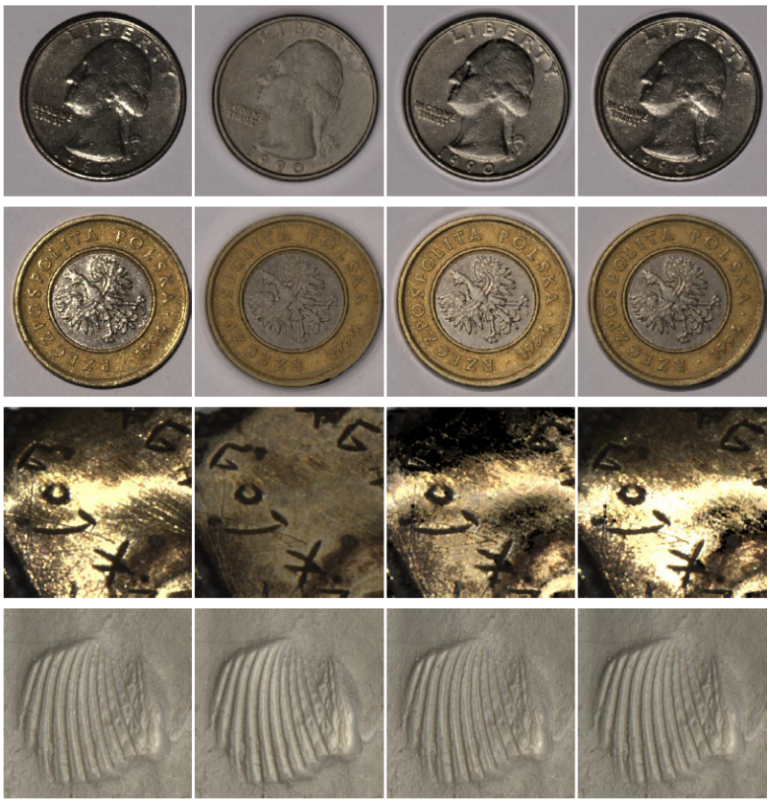


Figure 2.1: ***Direct interpolation.*** *For each row, here we show (from left to right): one of the original images in the MLIC; PTM-based rendering (section 2.3): HSH-based rendering (section 2.3); RBF rendering. RBF interpolation is capable of visualizing better novel virtual light directions, compared to classic methods.* Courtesy of Pintus et al. [105].

For this reason, it has been combined with compression approaches. Since the extraction of a set of basis has been demonstrated to work well to approximate the appearance [129], Ponchio et al. [116], in particular, proposes a relighting method based on a simple bilinear interpolation. In the pre-processing step, they create a novel relighting with RBF by using novel light directions that form a regular grid in the light direction space. Then, they combine this information with a per-pixel compression strategy based on Principal Component Analysis (PCA). At exploration time, they perform relighting from a novel light direction by applying bilinear interpolation to the pre-computed grid, and by selecting the final color from the compact PCA basis. Their data encoding is extremely suitable for online relighting, and a WebGL based interactive viewer has been realized that exploits the combination of those two techniques. Finally, another interesting thing is that PCA components can be stored in byte planes, and compressed as JPEG images to support compact relightable data exchange [116].

2.1.2 Visualization and Analysis

Direct exploration approaches are the least sensitive to calibration and raw data quality (e.g., for original data viewing without interpolation, the visual quality is simply the same as the capture sensor). For this reason, these techniques have been successfully employed in many application domains (e.g., Cultural Heritage - CH, Sec. 1.3), and they prove very suitable to directly treat complex materials with behavior that is difficult to model [47]. Original data visualization, and to a lesser extent interpolated data, is capable of reproducing both high-frequency components (e.g., highlights) of the reflectance field, and global effects, such as cast shadows and interreflections, that allow users to appreciate and understand the geometry at a meso/macro level. With an extremely dense acquisition, and a continuous light selection interface coupled with nearest-neighbor interpolation, the user can naturally select the original image to display, and interactively simulate controlled illumination. Direct exploration of the original images may be, however, difficult for visualizing massive image collections in remote settings. Moreover, the Graphical User Interface (GUI) for selecting a light is efficient only for regular acquisitions (e.g., domes), which offer a natural 2D parameterization. Further, the lack of continuity may make the perception of details difficult, as it is not possible, in most practical cases, to exploit motion parallax. Compression techniques combined with interpolation [116]) allow for simulating more flexible scenarios, including, for instance, moving a raking light, which is one of the classical methods used in real-world surface analysis and inspection. While the simulated continuous light motion may improve perception through motion parallax cues, it may also produce artifacts for cast shadows and highlights, especially if the sampling is not dense. Thus, some authors [47] have proposed to let users interactively tune the locality of the interpolation, but this approach is far from optimal.

2.2 Single-image data fusion

While direct exploration strives to provide direct access to all the individual images composing the MLIC, single-image fusion approaches have the goal to maximize the information that can be conveyed in a single static image. Often the performed data fusion produces physically inconsistent or impossible images but serves to increase the amount of shape information (less often appearance) presented in a single view of the object.

The most straightforward way, and classic way, to extract a single image that conveys object surface information from a MLIC is the albedo visualization. This is a light-independent view of the low-frequency (typically constant), spatially-varying BRDF across the surface [2, 24, 154]. Since the albedo removes highlights or other high-frequency effects that might disturb the underlying content in the visualization, this approach is often used for objects where pigment information is important (e.g., think of a painting inspection). However, a large amount of information about small, fine surface details is contained in the high-frequency part of that signal and is completely lost. Motivated by artistic techniques, which bring-out fine-scale details, while at the same time preserving the global object shape, several methods perform a multi-scale analysis of the MLIC signal at different spatial resolutions and frequencies, and then reconstruct an enhanced image that combines detailed information at each scale across all the input images. Fattal et al. [43] decompose each image in a MLIC into a series of detailed images plus a base image. The detailed

(a) *MLIC composed of three images*

(b) *Two enhanced results*

Figure 2.2: ***Single-image data fusion.*** *The input data consists in a 3 image MLIC (2.2a). The multi-scale approach of Fattal et al. [43] has been applied and two possible results have been obtained with different parameter configurations (2.2b). The enhanced view on the left locally highlights details and flattens the global appearance of the object, while the result on the right implicitly uses some shadows to increase detail perception.* Courtesy of Fattal et al. [43].

images are the differences between two bilaterally filtered versions at consecutive scales, while the base image encodes the low-frequency shading. The final visualization is built by a weighted sum of those contributions. Fig. 2.2 shows both an input 3 images MLIC and two possible results given different parameters, which drive the image fusion. One enhanced version makes the local finest details pop up, but flattens the global appearance of the object, while the other output implicitly uses some shadows (no shadow map has been explicitly extracted though) to augment the global perception of the object geometry. Since some high-frequency signals are generated by phenomena that are not related to local surface topology, or, in other terms, to real details or surface features (e.g., globally cast shad-

(a) *A car engine (original vs enhanced*

(b) *A flower plant (original vs enhanced*

Figure 2.3: ***Single-image data fusion for illustrative visualization.*** *Some MLIC-based fusion methods are capable of enhancing visual information in complex shapes, e.g., mechanical parts (2.3a), and of removing texture data through a digital visual abstraction that still preserves important features across the viewed surface ((a))(2.3b).* Courtesy of Raskar et al. [121].

ows), some algorithms integrate this approach with an additional step of shadow detection and removal [167]. Other works exploit shadows cast by multi-light acquisitions to classify spatial changes among those caused by spatially-varying materials across the surface, rather than depth discontinuities [121]. They use that information to build a single image visualization, which superimposes a stylized rendering on a processed version of the original input MLIC that removes unnecessary details while increasing the three-dimensional perception. This results in a final toon-like visualization, which highlights boundaries between the various shapes in the framed scene. Fig. 2.3 shows how this kind of illustrative technique has been employed to highlight the boundaries of complex shapes (Fig. 2.3a) and to remove texture data that contains less information through an abstraction strategy that preserves important features across the viewed surface. In this approach, an interactively controllable ”degree of abstraction” parameter controls the number of local details that can be removed (Fig. 2.3). When MLIC acquisition is performed also in a colored or even multi- or hyper-spectral framework [28], it is possible to exploit not only detailed information from the varying light position/direction but also from shading-independent

wavelength response across visible and invisible spectra. Once light-position-dependent components have been fused, the multi-spectral signals can be combined to produce false color visualizations, which fuse per-pixel contributions into a single image with highlighted details and main polychrome regions [147]. In this context, a common strategy to visualize high-dimensional chromatic data is to perform false-color PCA visualization. This data-reduction approach computed on a multi-spectral MLIC proved to be very powerful in significantly increasing the legibility of almost-faded details on the surface [146] and subsurface layers (e.g., underdrawings in paintings or manuscripts [150]), which are completely invisible in the original photographs.

2.2.1 Visualization and analysis

The combination, through a data fusion process, of all the MLIC information in a single image has the inherent disadvantage that some information is lost, interactive browsing is no more possible, and motion parallax cannot be exploited. On the other hand, the data-fusion process produces images that convey much more readable information than any single image in the MLIC. Compared to a photorealistic visualization by using direct exploration (Sec. 2.1), single image methods are capable of revealing both tiny surface texture elements and patterns, and global shape features (e.g., depth discontinuities in cluttered/complex environments), which are otherwise poorly visible or completely invisible in the photorealistic version or the original images. Since they try to emphasize geometrical features and to convey a visualization with a high level of abstraction, these techniques can provide useful insights into the nature of the real material appearance. For these depicting capabilities and their efficiency, these data fusion methods are employed for several tasks. Examples include visualization of surface texture and cracks in rock [43], fur visualization [43, 167], as well as venation and texture enhancement in botanic visual inspection [43, 121, 167]. When multi- or hyper-spectral capture is available, the shading-independent response to several different wavelengths can be used to enhance small poorly visible inscriptions (e.g., incisions on stones) [28, 146], to visualize abundance maps of pigments across the acquired object [28, 147, 150], or to improve the visibility of faded inks in old written text [147]; moreover geometrical structures and irregularities under the surface layer (e.g., pictorial layer visualization) can be made visible together with geometrical gradients. Moreover, these methods may produce a meaningful visualization even with a small amount of input data (e.g., $3-4$ images). For this reason, they have also been used for the enhancement of images during dynamic capture. Most of the approaches, moreover, do not require a camera and light calibration, since they perform the analysis just starting from the intensity changes in the image stack. The fact that they analyze multiple images at once makes them more effective than single-image enhancements or exaggerated shading techniques directly applied to 3D geometry [43]. They are employed to display scenes with low-contrast depth edges, and complex geometries such as mechanical parts (Fig. 2.3) or medical data (e.g., in endoscopy [121]). Besides, during the visualization, the user has only to tune a small number of parameters to interactively weight the contribution of various surface behavior components (e.g., weights of base and detail signals [43], amount of detail enhancement [167], or degree of abstraction). Single image approaches are also very suitable for storage and communication purposes in the context of web-based visualization platforms, since they exhibit a compact and compressed representation, and are very appropriate for printing. The limitation to static visualization may

be an interesting area of future work, since multi-image data fusion might be used in the context of real-time exploration methods, e.g., by fusing only subsets of images near the currently, interactively determined, focus light, or by exploiting a globally fused field in combination with other methods to improve shape and detail perception while performing relighting or other interactive tasks.

2.3 Relightable images

While single-image data-fusion mostly focuses on static visualization, offering the possibility of dynamically relighting the models is at the core of relightable model based approaches, which analyze the MLIC to learn the relation between measured images and incident illumination, and synthesize it in a relighting model. This set of methods, *relightable images* strives to directly model the reflectance field, without separating shape and material contribution. The techniques require, in addition to the MLIC, metadata with information on camera and light parameters to reason on the relation between measured reflectance and incoming light direction and intensity.

2.3.1 Polynomial Texture Maps

The seminal insight behind relightable images, introduced by Malzbender et al. [81] in terms of Polynomial Texture Maps (PTMs), was to replace classical multi-light processing based on Photometric Stereo [154] with a more flexible interpolation-based approach, which transforms the high amount of data in RTI image stacks into a compact multi-layered representation focused on approximating the measured appearance as a function of light direction. The method is based on a per-pixel bi-quadratic polynomial interpolation of the image stack producing a spatially varying reflectance I as a function of the light direction:

$$\begin{aligned} L\left(u,v,l_u,l_v\right) = a_0(u,v)l_u^2 + a_1(u,v)l_v^2 + a_2(u,v)l_u \\ + a_3(u,v)l_u + a_4(u,v)l_v + a_5(u,v) \end{aligned} \tag{2.2}$$

where l_u, l_v, are projections of the normalized light vector into the local texture coordinate system. Given n+1 images, the best fit at each pixel is computed using singular value decomposition (SVD) to solve the system of equations for $a_0 - a_5$:

$$\begin{bmatrix} l_{u0}^2 & l_{v0}^2 & l_{u0}l_{v0} & l_{u0} & l_{v0} & 1 \\ l_{u1}^2 & l_{v1}^2 & l_{u1}l_{v1} & l_{u1} & l_{v1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ l_{un}^2 & l_{vn}^2 & l_{un}l_{vn} & l_{un} & l_{vn} & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_5 \end{bmatrix} = \begin{bmatrix} L_0 \\ L_1 \\ \vdots \\ L_n \end{bmatrix} \tag{2.3}$$

At each pixel, given the six parameters $(a_0 - a_5)$ computed from data fitting, and a light direction $\vec{l}$, the reflectance is computed as:

$$I\left(\vec{l}\right) = a_0\vec{l_u}^2 + a_1\vec{l_v}^2 + a_2\vec{l_u}\vec{l_v} + a_3\vec{l_u} + a_4\vec{l_v} + a_5 \tag{2.4}$$

This method was the basis of the development of the RTI framework. A drawback of the quadratic model is that it captures only the low-frequency reflectance behavior, and it is not capable of reproducing high-frequency effects like specularities and shadows. The use of high order polynomials has been also tested [164], where third-order coefficients have been used, increasing the quality of the relighted images at the expense of increased oscillations in case of unreliable data. High-frequency effects may also act as outliers in the fitting process, damaging the accuracy of the reconstruction. For this reason, robust fitting methods(robust PTM [36], accelerated guided robust matte-model fitting [107]) have been successfully proposed. Robust PTM explicitly deals with highlights and shadows. This version lays its foundation on the assumption that the matte component is inlier whereas the specular and shadow components of the surfaces are outliers based on the solution which comes from robust regression, Least Median of Squares (LMS). It extended PTM in two ways: (1) First, the least-squares(LS) regression for the underlying matte model is replaced by a robust regression, the LMS method [124]. By utilizing the LMS, which generates outliers "on a silver platter", without any intervention the regression process robust to outliers. This means that only a majority of the n pixel values obtained at each pixel need be matte, with specularities and shadows automatically identified as outliers.; (2) Secondly, it furthers adds, an additional, interpolation level by modeling the part of in-sample pixel values that are not completely explained by the matte PTM model via a Radial Basis Function (RBF) interpolation. The RBF model takes care of features such as specularities and shadows that change abruptly with lighting direction. Robust PTM, start off by replacing the variable of polynomial function (equation 2.2) with $(l_u, l_v, l_w, l_u^2, l_{uv}, 1)$ where w $=\sqrt{1-u^2-v^2}$. Then it solves the model in a robust version, using LMS, with the upshot that both the shadows and specularities are identified automatically without the need for any thresholds.

Fornaro et al. [44] employ a PTM to model the matte signal and compute the distance between that matte function and the intensity of each pixel for all lighting directions. The average of those distances is considered a measure of the pixel glossiness and used to add a Phong based shading component to the rendering. This method is less general than the previous approaches since it makes the implicit assumption that the object is a reflective surface. Similarly, in the method proposed by Macdonald [78, 79] the matte model is used to extract the albedo and the normals (Fig. 2.5). Specularities are then obtained in the following way: first, the specular quotients are calculated as the ratio of the actual intensity divided by the Lambertian intensity, then a modified Lorentzian function is fitted on them (figure 2.5). The method, however, has not been used to interactively render images with arbitrary light directions.

2.3.2 Hemispherical Harmonics

Similar to the Fourier series, spherical harmonics are a good set of basis to describe functions on the surface of a sphere, and low order sets of those bases are typically used in modeling low-frequency reflectance, e.g., applied in photometric stereo scenarios [8]. In typical MLIC acquisition setups, the incident light at any surface point is defined on the upper hemisphere only, and full spherical representation is not needed. For this reason, the hemispherical basis defined from the shifted associated Legendre polynomials [41] has been introduced to represent image irradiance [88].

Hemispherical Harmonics(HSH) components are expressed as functions of the angles

for azimuth φ and co-latitude θ over the hemisphere in terms of l_u and l_v:

$$\varphi = (l_v, l_u) \;\; and \;\; \theta = acos\left(\sqrt{1 - {l_u}^2 - {l_u}^2}\right) \tag{2.5}$$

The basis functions $H_l^m(\theta, \varphi)$ which are orthogonal over the two-dimensional support $\left[0, \frac{\pi}{2}\right] \times [0, \pi]$ with respect to both l and m, are defined as:

$$H_l^m(\theta, \varphi) = \begin{cases} \sqrt{2}\widetilde{k_l^m} \cos(m\varphi)\widetilde{p_l^m}(\cos\theta), & \text{if } m > 0 \\ \sqrt{2}\widetilde{k_l^m} \sin(-m\varphi)\widetilde{p_l^m}(\cos\theta), & \text{if} m < 0 \\ \widetilde{k_l^0}\widetilde{p_l^0}(\cos\theta), & \text{if} m = 0 \end{cases} \tag{2.6}$$

where $\widetilde{p_l^m}$ are the associated Legendre Polynomials p_l^m shifted from the sphere to the hemisphere and k_l^m are the corresponding normalization factors.

The per-pixel reflectance is computed as:

$$I\left(\vec{l}\right) = \sum_{l=0}^{n-1} \sum_{m=-l}^{l} \alpha_l^m H_l^m\left(\vec{l}\right) \tag{2.7}$$

where H_l^m are the hemispherical harmonic (HSH) basis functions defined in equation 2.6.

To create relightable images from a captured dataset, the projection of the images onto the elements of the basis up to a maximal order is estimated. Typically, the maximal order used is 1 (4 coefficients), 2 (9 coefficients), or 3 (16 coefficients). It has been proved that good results are obtained with sixteen coefficients per-pixel [77], which consist of four first-order, five second-order and seven third-order terms.

Relightable images encoded with HSH coefficients are also currently used in different application domains and are supported by RTI viewers (section 2.6). HSH fitting is generally preferred to PTM since it captures better high-frequency reflectance behaviors. However, the analysis performed by Zhang and Drew [164] shows that, with a similar number of coefficients, polynomial and HSH coefficients provide similar results.

2.3.3 Discrete Modal Decomposition

Discrete Modal Decomposition (DMD) has also been proposed to generate a continuous model of the local reflection. The DMD method is based on a projection on the modal bases, which are composed of elementary forms that take their origin from a structural dynamic problem. Pitard et al. [111, 112] adopt this technique for modeling angular components of the local reflectance by devising new modal shapes called Reflectance Modal Basis (RMB) from the set of images based on the Discrete Modal decomposition. They model the surface reflectance function, f, as the sum of a linear combination of the modal vectors and the residue of the decomposition R_n, which can be expressed as:

$$f(\theta_v, \varphi_u)(\theta_i, \varphi_i) = \sum_{k=1}^{n} \lambda_k\left(\theta_v, \varphi_u\right) \mathtt{Q}_k(\theta_i, \varphi_i) + R_n \tag{2.8}$$

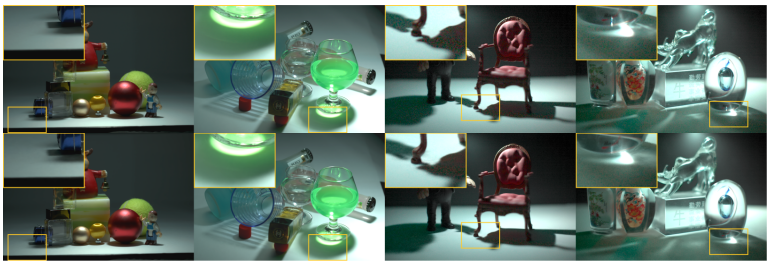


Figure 2.4: ***Light Transport Matrix Reconstruction.*** *The first row shows a series of ground truth images of a light transport matrix column which is not included in the input data used to produce the relightable image model. The second row is the same matrix columns reconstructed by using a trained relighting Neural Network.* Courtesy of Ren et al. [123].

where the modal coefficients λ_k can be obtained as the projection of the vector of measured luminance $\mathbf{L}$ onto each mode ($\mathtt{Q}_k$, k=1,...n, where n is the number of modes) of the non-orthonormal basis.

A comparative analysis with other fitting techniques has shown that the DMD is well suited for approximating the complex physical behavior of light reflections, especially for shiny reflective surfaces [111, 112].

2.3.4 Deep learning approaches

An (artificial)neural network (aka deep learning) is a powerful learning algorithm inspired by how the brain works. Convolutional Neural Network (CNN) is a variant of Neural Network that lays its foundation, mainly, on the convolution operation. Over the past few years, this method has been successfully adopted to solve problems related to appearance measurement and analysis (to estimate reflectance map and to create relightable images) and an effective result has been obtained [59, 122, 123, 162].
In contrast to the methods that recover the relighted image independently per pixel CNN based global methods have been also proposed. Unlike the methods mentioned above, this method considered the general problem of image relighting and not classical RTI setting.

Xu et al. [162] have proposed a method to learn the behavior for entire image patches using specialized Convolutional Neural Networks (CNNs). Two networks are trained, one to select optimal light sampling (that, therefore, is then fixed with a certain tolerance), and the other to relight the scene given a novel light direction. The method can reasonably model global effects (e.g., shadows and interreflections) and complex materials (general isotropic) with an extremely small number of images compared to state-of-the-art techniques (e.g., only 5 images). However, the rendering of global high-frequency behaviors, such as sharp cast shadow boundaries, remains a challenging issue. That is the reason why that approach is more suitable (and can obtain good error metrics) when environmental light is provided rather than a directional light.

Another deep learning relighting method is the one proposed by ren et al. [123]. This method bypass reconstruction by directly modeling the scene's light transport function. Assuming distant illumination, the light transport function, also called basis function,

$T(x, \omega)$, maps incident illumination from direction ω to outgoing radiance at pixel x (towards the camera), and allows for the scene to be rendered under novel distant lighting as:

$$I(x) = \int_{\mathcal{L}} T(x, \omega) L(\omega) d_{\omega}, \tag{2.9}$$

where $L(\omega)$ is the radiance of the incident illumination from direction ω.

The light transport function can be sampled by capturing images under different lighting conditions; for example, an image of the scene under a single directional light from direction ω_j, yields the sample:$I_j(\cdot) = T(\cdot, \omega_j)$. Image-based relighting methods use a set of such samples, $(I_j, \omega_j) \mid j = 1, 2, ..., k$, to reproduce scene appearance, I_n, under a novel light, ω_n. In the deep learning relighting approach, aka Relight-Net, light transport is formulated as discrete samples of continuous light transport function and estimated using neural networks.

Given a set of k images of the scene, $I_1, I_2, ...I_k$, captured under predefined directional lights, $\omega_1, \omega_2, ..., \omega_k$ respectively the reconstructed image, I_n, that would be produced by a novel directional light, ω_n, via a relighting function, $\phi(\cdot)$ is:

$$\begin{aligned} I_n &= \phi(\omega_n; I_1, \omega_1; I_2, \omega_2; ...; I_k, \omega_k) \\ &= \phi(\omega_n, S_1, ..., S_k) \end{aligned} \tag{2.10}$$

where $\mathcal{S} = S_j = (I_j, \omega_j)$ and ω_j is a 2D light direction. usually, S_j is a 5-channel input per-sample.

Xu et al.[162] modeled the relighting function, $\phi(\cdot)$, as a deep convolutional neural network (CNN) referred as Relight-Net. The Relight-Net trained using a large synthetic dataset consisting of procedurally generated shapes rendered with complex spatially-varying BRDFs and demonstrated that it can reconstruct high-frequency light transport effects like specularities and cast shadows.

2.3.5 Visualization and analysis

This class of relightable models is very general and easy to apply. By avoiding to solve the complex problems of separating shape from material contributions, and of modeling complex light transports, these methods can, in principle, provide visually realistic approximations while interactively controlling a virtual light. Similar to the local interpolation methods (section 2.1), this class of fitting techniques can be used to simulate interactive lighting or raking light. The user is provided with a simple interface element to select a virtual light source (typically a Gaussian sphere projected in 2D), and the visualization algorithm uses the parameters of the fitted analytic model to transform that light direction input in the proper relighted image; besides, the user might pan and zoom, and select regions of interest within the relightable image. In addition to simple interpolation, moreover, more complex modifications can be made to the illumination, including spatially-varying intensity and direction control.

Due to the rich viewing experience that this approach provides, this is the most used approach in a wide range of visualization and analysis tasks. Interactive relighting controls allow viewers to better understand object morphology in terms of meso- (e.g., local surface bumps, roughness, superficial patterns) and macro-structures (e.g., global curvature,

convexity and concavity cues) by using an interface that mimics the typical real-world inspection (e.g., with a raking light). This fact makes them a good virtual alternative to real-world inspection under controlled light for surface quality control in many fields from CH study and inspection, criminal forensics, art conservation, and paleontology [88], to micro-mechanics, biomedicine, and horology.

In most cases, however, only the low-frequency behavior is sufficiently accurate for typical acquisition sampling rates (i.e., hundreds of images). A full model in terms of geometry and material is more amenable to produce physically-valid relighting; such methods are only computable for specific classes of objects, but relightable images are more generally applicable and have been shown to help in understanding how different optical features as matte nature or glossiness, and generally heterogeneous materials, are distributed across the object, as well as to visually highlight surface grain/granularity, damages, weathering patterns, erosion, and general decay.

All the various methods, based on attempting to directly map, through a small set of parameters, the changes in perceived color to change in illumination direction, tend to exhibit visual artifacts due to the global interpolation in the light space. To overcome this issue, LTM-based methods and CNN-based methods strive to learn more information on global transport. In principle, LTM-based methods are capable of producing an extremely accurate relighting, with anisotropic, transparent and highly complex local behavior of materials and global illumination effects (figure 2.4). Unfortunately, while they meet these goals, LTM approaches pose several problems in terms of capture complexity, and memory occupancy for both storage and communication issues. In terms of capture, to avoid hallucinations, a very high-frequency MLIC should be employed. Moreover, the size of the typical reconstructed light transport function for each pixel is about 200k samples [142], so, even for small images, the representation tends to pose very hard problems in terms of storage, transmission, and computation required to meet interactive rates during relighting. The use of neural networks [123], trained for each specific dataset, tries to cope with this problem, by producing a relatively small model with a tractable memory footprint and fast relighting performance. However, the methods proposed so far tend to produce hallucinations, and are, therefore, not very appropriate to visual inspection tasks.

For these reasons, the most employed methods in practice are still those based on low-frequency fitting (e.g., PTM, HSH). These representations are easy to compute and to compress, eventually exploiting available image-based encoding (e.g. JPEG) to compress the various layers, leading to straightforward incorporation in web tools and mobile devices (section 2.6). In these tools, visualization is often integrated with more complex knowledge organization and representation software, and within systems for the annotation of specific sets of objects, e.g., in the arts and humanities fields. Some of those tools allow the user to link comments and annotate the relightable images during the visualization experience [44], and export or store this information as metadata. This opens the way for collaborative visualization environments.

2.4 Non-photorealistic enhancements

Relightable image models are often combined with illustrative techniques, to improve the understanding of scenes beyond what is possible by sticking to photorealistic approaches. Differently from static methods, which extract single images that convey an enhanced view

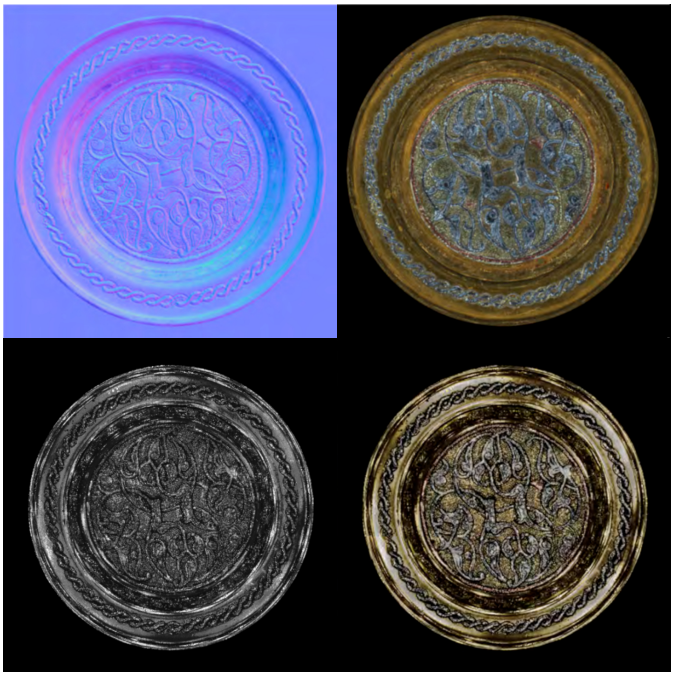


Figure 2.5: ***Layered models.*** *This class of models combines a matte model with a specular/shadow layer to better render complex materials. In this figure, we show several layers that are modeled from MLIC data: (1^{st} row, left to right) surface normals, matte albedo; (2^{nd} row, left to right), the specular quotient (the ratio between specular and diffuse intensities), and the specular color.* Courtesy of MacDonald [79].

of the object (already seen in section 2.2), the methods described here focus on interactive-rate solutions that modify the employed representations, or that post-process the generated images.

The most popular strategies to improve surface detail readability in real-time relighting applications are the *Diffuse Gain* [15, 40, 54, 130, 145] and the *Specular enhancement* approaches [6, 15, 40]. The *Diffuse Gain* has been presented for the first time by Malzbender et al. [81]. The main idea is to apply a transformation to the fitting parameters of the matte model function (in their case, the six PTM parameters). By non-linearly changing those luminance coefficients, they are capable of increasing the visual perception of the curvature of the surface response to light direction variations. That transformation is designed to maintain the same per-pixel normal field. This reflectance-based contrast improvement, together with the dynamically moving light position, proved to be an efficient and fast way for aided morphology inspection and the visualization of surface subtleties not discernible to the naked eye [89, 130, 145]. Having an underlying normal vector field, all types of materials can be visualized by adding a specular component to their analytical or data-driven BRDF function. This is an old and widely used technique, due to its simplicity, efficiency and effectiveness in enhancing meso and micro structures across the object surface [87, 89], to better discern object content [91, 130, 145]; figure 2.6 is an example

of how those exaggerated shading can reveal not only shape content but also semantic information about how the object has been manufactured [146]. The most simple specular model consists in a combination of a specular color k_s and a non-linear (typically exponential) function driven by a roughness parameter α. During the visualization, the user can easily tune those parameters in real-time [80] while dynamically changing the light source to even extreme conditions, up to raking light [67], to obtain the desired effect and to highlight different shape properties and features [54]. The *Specular Enhancement* can be used together with the color information blended with it [53], or in a visualization that removes the albedo or generally the chromatic information, by displaying only the reflectivity component, and by focusing only on the amplification of slope perception [28, 145].

Besides those two ways of enhancing relightable image visualizations, which are of paramount importance for visual details inspection, other techniques are less used, although equally effective for the task. One consists in applying unsharp masking to the luminance field [15, 130]. Others exploit a light that is not constrained to have a unit norm, and goes beyond that physically acceptable condition, and perform light extrapolation [54, 95]. This non-photorealistic representation provides a more "oblique", raking light, and higher image contrast. Of course, by enabling virtual relighting, one can also increase the number of virtual lights. This can be used to increase object readability by automatically choosing the appropriate set of lights that augments details perception. Palma et al. [98] present a multi-lighting detail enhancement that, given a single light direction selected by the user, computes a virtual lighting environment that aims at increasing the local contrast across the image. This is a non-photorealistic method because that virtual lighting is not feasible and reproducible in the real world due to the high number of lights, the very localized contribution (or support volume) of each light (which is not physically plausible), and the real-world interreflections which would ruin the original enhancement purpose by smoothing out the desired sharpening effect.

To enrich the quality of the visualization enhancement, the aforementioned interactive techniques are typically combined with the real-time normal enhancement mode [80] to increase granularity and sharpness in the final visualization. Other approaches consist in directly applying the same rationale of unsharp masking [98], or a simple contrast enhancement operator [47], to the normal vector field.

2.4.1 Visualization and analysis

With the same purpose as the methods seen in section 2.2, non-photorealistic enhancement is aimed at producing a visualization that increases details perception. Similar to single-image data-fusion methods, many of the approaches are specially devoted to highlighting microfeatures (e.g., roughness) and mesostructures/patterns across the object surface, and mostly focus on shape rather than material enhancement. Unlike single-image data-fusion methods, however, most of the illustrative techniques are interactive and tightly coupled with both a dynamic relighting, and, in some cases (e.g., *Diffuse Gain*, *Specular enhancement*, and *Unsharp Masking*) even with real-time MLIC acquisition and enhancement [80]. Direct interactive control of parameters is generally used to tune various contributions of different appearance layers. Since most of the methods can be applied to the data stored in classic relightable image formats, the same advantages in terms of portability, compression and communication described in section 2.3 are valid here as well.

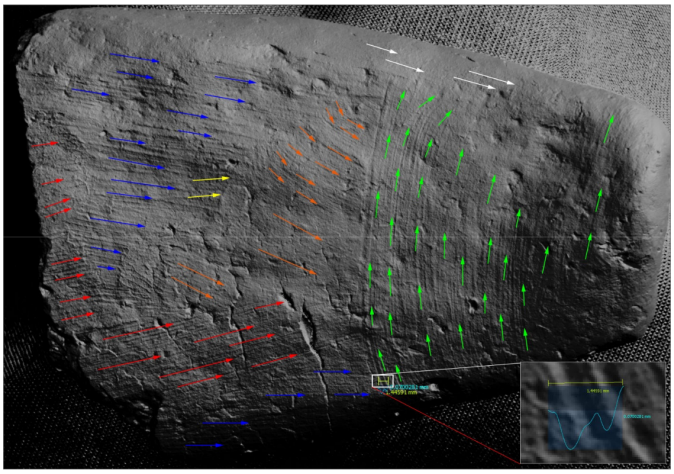


Figure 2.6: ***Non-photorealistic enhancement.*** *Exaggerated shading helps in the visualization of the finest object details. Here, for instance, directions of smoothing marks across the surface have been highlighted by this non-photorealistic way of display MLIC-derived data.* Courtesy of Van et al. [146].

Many of the non-photorealistic techniques focus more on the shape than on the appearance. For this reason, many of them rely on the computation of normals from MLIC to convey shape information, which does not apply to all imaged scenes. When this is possible, those non-photorealistic enhancements have proved to be successful in highlighting hardly visible features (e.g., poorly visible inscriptions [53]). In some cases, to highlight geometrical details, the appearance is even changed on purpose, e.g., when applying specular enhancement on completely matte objects to emphasize surface finish, so optical material interpretation is not applicable and misleading. Further, some visualization techniques consider the color or appearance information as a cluttering element, and they remove that to bring the wanted details to light [91]. This is true also in the cases where the appearance/color information is not removed but rather exploited; the multi-spectral signal is used to produce false-color images that, rather than containing information about the optical behavior of the surface, make it visible underlying shapes and features [147]. Since the scope of non-photorealistic methods is the enhanced visualization of semantic structures and details across the surface, visualization is not so dependent on an accurate light/camera calibration, since, generally, small scale calibration errors do not change much the overall information content of the imaged scene. Unlike the methods in section 2.3, illustrative techniques are thus more robust to acquisition noise and inaccurate, relaxed computation of light, camera, and radiometric parameters. Due to the combination of a compact data format, the user-friendly interface, the robustness to noise or poor calibration, and the effectiveness of those illustrative techniques, this group of methods are the most used ones in a large number of daily applications (section 1.3).

2.5 Normal estimation

Photometric Stereo(PS) is a shape reconstruction technique that relies on multi-light image collections. It is a technique targeted for estimating the surface normal, surface orientation, of an object given MLICs, a set of images relighted from different directions of illuminations. It was first introduced by Woodham [154] and assumes only a Lambertian surface. Since then, a lot of work targeted to extend the original idea to surfaces with general reflectance properties and improving the precision has been done. In recent years, in addition to the classical techniques, (deep)neural network-based photometric stereo approach is being emerged [17, 18, 57, 128, 141].

2.5.1 Photometric Stereo(PS) approaches

Most of the existing methods for photometric stereo [101, 131] assume a simplified reflectance model, such as the Lambertian model for its simplicity. However, this assumption doesn't work since most of the real-world objects are non-Lambertian. Thus, many photometric stereo algorithms have been developed to deal with non-Lambertian materials. The most popular ones are based on outlier rejection and the use of the Lambertian model for the remaining inliers. Within this category, various methods have been proposed that rely on different principles, such as RANSAC [90], median values [85], expectation-maximization [156], sparse Bayesian regression [62], Least Median of Squares [36, 107], Low-rank matrix completion and recovery [155], or Sparse Regression [63].

Unlike outlier rejection methods, methods based on a sophisticated (analytical) reflectance model fit a model to all observations. This is achieved by solving complex optimization problems. Many sophisticated analytical reflectance models have been proposed to approximate the behavior of non-Lambertian materials, including the Torrance-Sparrow model [46], Ward model [22], Cook-Torrance model [125], etc. The downside of these types of methods is that they can only handle limited classes of material. On the other hand, example based methods typically require an example object with known surface normal, shape and reflectance, to be placed in the scene. Usually, this requirement limits its practical use, so other approaches employ a dictionary of BRDFs to render virtual examples that guide the normal estimation problem. To this end, Hui and Sankaranarayanan [58] proposed a BRDF dictionary to render virtual spheres without using a real example object. Learning based methods estimate mappings from measured intensities under known (or unknown) lighting to surface normals by using machine learning tools or deep learning techniques[17, 18, 61]. Deep learning is a powerful learning method inspired by how the brain works. Convolutional or Fully-Connected Neural Network based methods have recently replaced traditional PS techniques. To mention, Santo et al. [128] proposed a deep fully-connected neural network, called Deep Photometric Stereo Network (DPSN), to learn the mapping between reflectance observations and surface normals in a per-pixel manner given a fixed number of observations captured under a pre-defined set of light directions. In this work, for each image point of the object, all its observations and light directions are concatenated to form a fixed-length vector, which is fed into a fully-connected network to regress a single normal vector. The weakness of this work arises from the assumption that the light directions are pre-defined and remain the same between training and prediction phases, which in turn limits its practical use. Chen et al. proposed two

different methods based on convolutional neural networks [17, 18]. The first method [18] proposes a flexible fully convolutional network, called PS-FCN, for estimating a normal map of an object. The network consists of three components, namely a shared-weight feature extractor for extracting feature representations from the input images, a fusion layer for aggregating features from multiple input images, and a normal regression network for inferring the normal map. The second method proposes a two-stage model named Self-calibrating Deep Photometric Stereo Networks (SDPS-Net). The first stage of SDPS-Net, denoted as Lighting Calibration Network (LCNet), takes an arbitrary number of images as input and estimates their corresponding light directions and intensities. The second stage denoted as Normal Estimation Network (NENet), estimates a surface normal map of a scene based on the lighting conditions estimated by LCNet and the input images. In both cases, to simulate real-world, complex non-Lambertian surfaces they trained their model on synthetic datasets created using shapes from the blobby shape dataset [64] and the sculpture shape dataset [152], and BRDFs from the MERL BRDF dataset [84]. Another Neural Network based method, which has demonstrated the possibility of recovering normals better than traditional methods, is the one proposed by Ikehata [60]. This network accepts an arbitrary number of input images, merges them into the intermediate representation called observation map which has a fixed shape, and then regresses the normal map. This network can directly learn the relationships between the photometric stereo input and surface normals of a scene. For a detailed and up-to-date survey please refer to Shi et al. [132].

2.6 Interactive visual inspection tools

Several interactive MLIC-based visualization tools have been described in the literature, and a few of them have been made publicly available and are actively maintained. Some of them are generically applied to datasets captured with different cameras and lighting setups, while others are used to visualize only data acquired by a specific acquisition device (e.g., a light dome) and/or pre-computed versions of the relighted image [79, 147]. While some of them allow for visualization on the web or portable devices [97, 116], others are developed considering only desktop applications with fully resident data. A summary of the major available tools is shown in Table. 2.1.

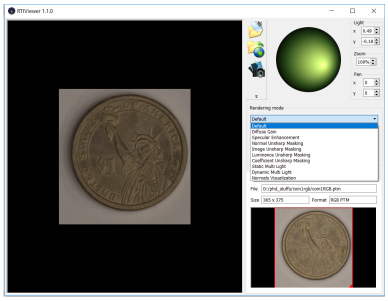
The most popular application is *RTI viewer* (figure 2.7a). It has been primarily developed by the ISTI/CNR Visual Computing Laboratory and is financed and distributed now by Cultural Heritage Imaging [20]. Complete credits and executable codes can be found at the CHI website. This tool allows for the visualization of relightable image files stored on the local file system and remote files through HTTP. It features a visible virtual torch, which allows the user to perform photorealistic image relighting based on PTM or HSH (section 2.3), and includes several non-photorealistic enhancements (section 2.4), including specular enhancement and image unsharp masking. Due to its free availability and ease of use, together with the CHI [20] capture software, it is widely utilized in different areas of a study presented in section 1.3. Another desktop application tool, but mainly used in image processing research works, is *APTool* [49] (figure. 2.7b). This tool allows for interactive exploration of raw MLIC data using RBF interpolation; it also creates relightable images by using PTM or HSH and extracts and visualizes feature maps. It can also interactively show per-pixel interpolated reflectance maps. This tool has been used in

Tool	Access	Visualization	Platform
RTIViewer [20]	Public	Photorealistic relighting (PTM, HSH); Specular enhancement; Diffuse gain; Normal, Image & coeff unsharp mask	Desktop
APTool [49]	Private	Photorealistic relighting (PTM, HSH, RBF); Normal/Albedo Maps; Appearance profile vis.	Desktop
PLDViewer [69]	Private	Photorealistic relighting; (Exaggerated)Shading; Curvature; Sketch	Desktop
ImageViewer [79]	Private	Original images; Outline, albedo and normal map; Specular components; Low-frequency fitting (PTM, HSH)	Desktop
WebRTIViewer [96]	Public	Photorealistic relighting (PTM, HSH);	Web
Relight-viewer [117]	Public	Photorealistic relighting (PTM, HSH, RBF, Bilinear)	Web
PLDWebViewer [69]	Public	Photorealistic relighting; (Exaggerated) Shading; Curvature; Sketch	Web
DigitalMateriality Viewer [32]	Public	Photorealistic relighting; Specular enhancements	Web

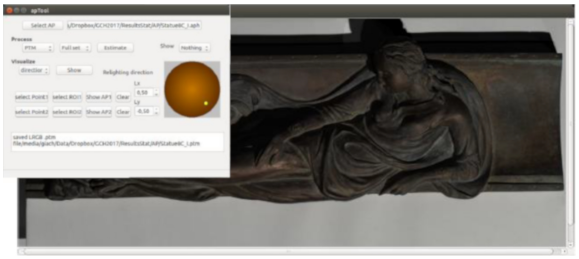
Table 2.1: ***Tools.*** *Major MLIC-based visualization software tools described in the literature.*

CH for the study of various objects such as coins, rock art, statues, and paintings [24, 47]. A multi-functionality application tool called PLDviewer (Portable Light Dome viewer) has been developed at KU Leuven [69], based on a full processing framework [147]. This tool accepts as input proprietary compressed file generated with a specific dome device and an undisclosed processing pipeline. It provides interactive relighting (photorealistic visualization), 3D model construction, and non-photorealistic visualization. The first can be performed with the help of two virtual light sources with tunable intensity (figure 2.7f). The viewer supports several rendering modes, e.g. albedo and ambient, and allows for multi-spectral visualizations in five spectral bands (i.e., Infrared, Red, Green, Blue, and UltraViolet). Six views of an object are shown in the interface for full 3D object creation, and the user can interactively visualize one at a time by switching between different views using arrow-keys. In addition, this tool allows for performing non-photorealistic enhancement (section 2.4) such as curvature coloring, exaggerated shading, and sketching. Furthermore, it allows for displaying the map of the response to the set of lights on single pixels (reflection maps) and local histograms. This tool is mainly developed for CH applications but can be extended to other application domains. The desktop visualization tool presented by MacDonald [79] (ImageViewer, figure 2.7e) is capable of displaying original photographic images and different kinds of modified renderings with the same lighting. Unlike RTIViewer, APTool, and PLDViewer this tool does not support interactive relighting. Several tools have been proposed also for web-based visualization of relightable images. WebRTIViewer [96] is a web-based multi-light visualization tool that allows for the visualization of the relightable images created using PTM or HSH in a standard web page, without the installation of any plugin and with a recent web-browser. PLDWebviewer is the web-based counterpart of the PLDViewer with similar functionality. Other web-based tools have been developed by the University of Basel [32]. In particular, the Digital Materiality Viewer (figure 2.7d) allows for interactive relighting and glossy

enhancement, as described by Fornaro et al. [44]. Relight [117] is a library to create and view (on the web) relightable images (RTI), which supports the encodings described by Ponchio et al. [116]. The library can be used to create relightable images based on PTM, HSH, and RBF, and exploits different compression and color-coding options. The visualization methods feature both an OpenGL Javascript library for RTI rendering on the web, which can be used to create web pages with single or multiple relightable images, and the Relight Viewer, a multiresolution viewer with a toolbar controlling a few basic options (figure 2.7c).

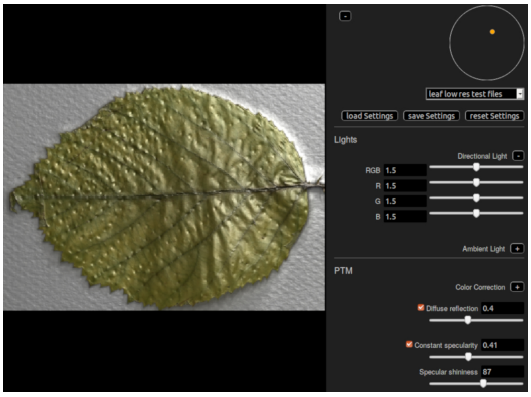


(a) *RTIViewer*



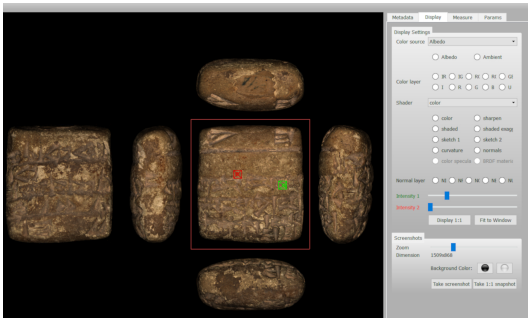
(b) *apTool*

(c) *Relight Viewer*



(d) *Digital Materiality Viewer*

(e) *ImageViewer*



(f) *PLDViewer*

Figure 2.7: ***Screenshots of visualization tools:*** *(2.7a) RTIViewer [20], the most popular solution for relightable image visualization; (2.7b) apTool, a research tool developed by University of Verona [49]; (2.7c) Relight Viewer, a tool part of a powerful library developed at ISTI/CNR Pisa [117]; (2.7d) Digital Materiality Viewer application developed at the Digital Humanity Labs (University of Basel) [32]; (2.7e) GUI of desktop visualization tools described by MacDonald [79] (this tool does not allow interactive relighting); (2.7f) PLDViewer GUI, which handles dense acquisition realized with the Portable Light dome [69].*

CHAPTER 3

Novel Multi-Light Image Collections for RTI and Photometric Stereo benchmarking

One of the challenge in the Multi-Light Image Collections (MLIC) community, specifically that relies on single view and wavelength MLIC data, is the lack of benchmarks to evaluate data processing tools on the kind of surfaces typically captured in real-world applications (e.g. quasi-planar surfaces made of heterogeneous materials with a wide range of metallic and specular behaviors). Relighting algorithms are typically tested on a few homemade acquisitions not publicly available. Whereas photometric stereo algorithm (recovering surface normals), nowadays, widely, tested on the only available public benchmark, DiLiGent, [132]. This benchmark could be used as well to test relighting quality. However, the images included are not representative of those typically captured in the real-world applications of MLIC acquisition.

Relightable images and Photometric Stereo algorithms are used mainly to analyze surfaces with limited depth, possibly made of challenging and heterogeneous materials. The performances of the different methods proposed are likely to vary with the shape complexity and the material properties. It would be, therefore, useful, to evaluate the algorithms on images of this kind, possibly with accurate lighting and controlled features.

In this chapter, we first present three novel MLIC benchmark datasets: one composed of acquired images (RealRTI) and two composed of synthetic images (SyntRTI and SynthPS) generated with Physically-Based Rendering techniques (PBR).

RealRTI is generated using 12 real cultural heritage objects, while SyntRTI and SynthPS are created from scanned 3D models of 5 cultural heritage objects with different

materials and materials distributions assigned. These datasets can be used for validation of MLIC-related algorithms (for instance: Photometric Stereo, Reflectance Transformation Imaging (RTI) relighting, material classification). SynthRTI and RealRTI have been used to evaluate the quality of neural network-based RTI method NeuralRTI, described in Chapter 5 [38]. SynthPS have been used for the evaluation of photometric stereo algorithms for cultural heritage usage, described in chapter 4.

3.1 MLIC benchmarks dataset

3.1.1 RealRTI

RealRTI is a dataset made of real MLIC acquisitions, made with devices and protocols typically used in the MLIC acquisition (see section 1.1). This dataset is composed of 12 multi-light image collections (cropped and resized to allow a fast processing/evaluation) acquired with light domes(using a setup of Ciortan et al. [24] and Pintus et al. [105]) or handheld RTI protocols(using a setup of Giachetti et al. [48]) on surfaces with different shape and material complexity. The items imaged are: (1) a wooden painted door (handheld acquisition, 60 light directions), (2) a fresco (dome acquisition, 47 lights), (3,4) two painted icons (handheld 63 and 72 lights), (5,6) two paintings on canvas (handheld, 49 lights and dome, 48 lights), impressions on the plaster of a leaf (7) and a shell (8) (light dome, 48 lights), (9,10) two coins (both with a light dome, 48 lights), (11,12) two metallic statues (dome, 48 lights and handheld, 54 lights). This set allows testing relight on different materials (matte, specular, metallic) and shapes with different geometric complexity(see figure 3.1).

The RealRTI dataset can be accessed at the following URL: `https://github.com/Univr-RTI/RealRTI`.

3.1.2 SynthRTI

SynthRTI is a collection of 51 multi-light image collections simulated by using the Blender Cycles rendering engine. It is divided into two subsets: SingleMaterial, featuring 24 captures of three surfaces with 8 different materials assigned (figure 3.2), and MultiMaterial, with 27 captures of the same three surfaces with 9 material combinations, as shown in figure 4.8. Each collection is subdivided into two sets of images, corresponding to two sets of light directions. The first is called Dome and corresponds to a classical multi-ring light dome solution with 49 directional lights placed in concentric rings in the l_x, l_y plane at 5 different elevation values (10,30,50,70,90 degrees). The second is called Test and includes 20 light directions at 4 intermediate elevation values (20,40,60,80). The idea is to use the classic dome acquisition to create the relightable images and evaluate the quality of the results using the test dataset. Figure 3.3 shows the Blender viewport with the simulated dome (left) and the light directions of the Dome and Test acquisition setups, represented in the l_x, l_y plane

The three geometric models used to build the dataset are bas-reliefs with different depths. The first is a nearly-flat surface, actually the 3D scan of an oil on canvas painting by W. Turner, performed by R.M. Navarro and found on SketchFab (*https://sketchfab.com*). The second is the scan of a cuneiform tablet from Colgate University. The third is the scan

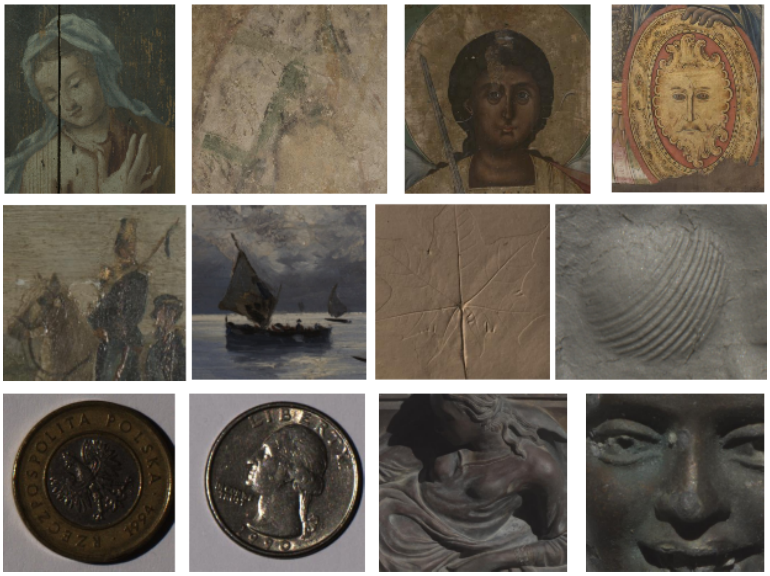


Figure 3.1: *Sample images of the RealRTI dataset representing the 12 different surfaces captured.*

Figure 3.2: *Sample images of the bas-relief object with the 8 materials assigned.*

of relief in marble "The dance of the Muses on Helicon" by G. C. Freund, digitized by G. Marchal. All the models are distributed under the Creative Commons 4.0 license.

For the single-material collections, we used 8 sets of Cycles PBR parameters, to simulate matte, plastic and metallic behaviors and material with subsurface scattering. These sets are reported in Table 3.1. We rendered small images (320×320) with 8 bits depth. The MultiMaterial subset is created assigning different tints and material properties to different regions of the three geometric models. We created 9 material combinations from 8 base diffuse/specular/metallic behaviors (Table 3.2) and 16 tints. The use of multiple materials may create problems in algorithms using global stats of the sampled reflectance

#	material	base color	metal	spec	rough	subs. R
1	matte white	0.8,0.8,0.8	0	0	0	
2	plastic	0.8,0.8,0.8	0	1	0.4	
3	plastic rough	0.8,0.8,0.8	0	1	0.5	
4	plastic rough2	.25,.25,.25	0	1	0.6	
5	plastic black	.25,.25,.25	0	1	0.4	
6	metal smooth	0.8,0.8,0.8	1	0	0.5	
7	metal rough	0.8,0.8,0.8	1	0	0.7	
8	subsurface	0.5,0.5,0.5	0	1	0.5	1,0.2,0.1

Table 3.1: *Parameters of the Cycles principled BSDF model used to create the 8 materials of the SynthRTI, single material dataset.*

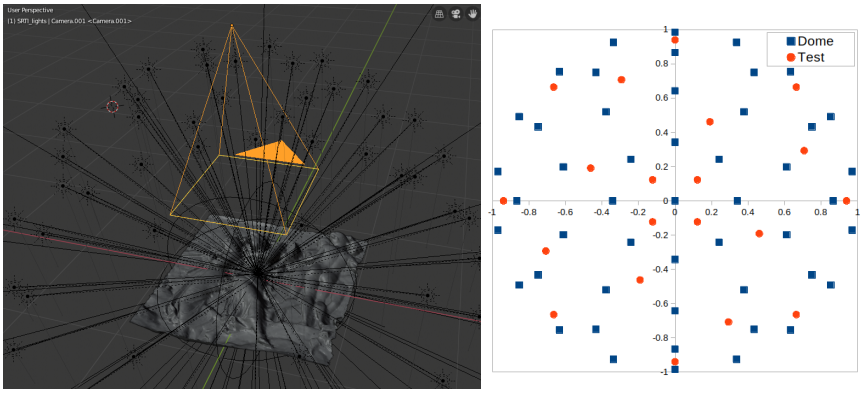


Figure 3.3: *Left, the virtual camera/lights configuration in the Blender viewport. We used directional (sun) lights. Right, the distribution of the light directions in the simulated dome subset (blue squares) and for the test evaluation (red dots).*

	r1	r2	r3	r4	r5	r6	r7	r8
metal	0	1	1	1	0	0	0	0
roughness	1	0.4	0.5	0.6	0.5	0.6	0.7	0.8
specularity	0	1	1	1	1	0	0	1

Table 3.2: *The combination of metallic, specularity and roughness indexes used in the SynthRTI, MultiMaterial dataset.*

to create mappings for dimensionality reduction.

SynthRTI can be exploited not only to test RTI relighting approaches but also to evaluate Photometric Stereo methods, as the rendering engine can output normal maps, as well as shadow and specularity maps for each image (see Figure 3.5). Ground truth normal maps and shadow maps are publicly released with the SynthRTI distribution, which can be accessed at the following URL: `https://github.com/Univr-RTI/SynthRTI`.

3.1.3 SynthPS: image collections and tasks

SynthPS is a set of multi-light image collections synthetically rendered with the physically-based reflectance models of the Blender Cycles engine. Each collection is rendered with 77 directional lights placed in concentric rings in the l_x, l_y plane at 9 different elevation values (10,20,30,40,50,60,70,80,90 degrees). Figure 3.9 shows the full set of light directions used. Lights are exactly directional and no ambient light is employed. The camera model is orthographic, the size of the rendered images is 320×320, and the depth is 16 bits linearly mapped. SynthPS is composed of two subsets: Single-material, i.e., all the items are created using constant materials; MultiMaterial, i.e., all the items (different from those used in the Single-material case) are textured with captured spatially-varying albedo and subdivided in patches with different roughness.

The Single material dataset has been created using the same geometric models used to create SynthRTI(Figure 3.6, top). For these renderings, we have set an orthographic camera looking at the object surfaces, removed the ambient illumination, and rendered the set of images with the 77 directional lights. For each model, we created 9 collections of images assigning 9 different uniform materials to the surfaces. The assigned materials simulate matte, plastic and metallic behaviors with varying gray achromatic albedo and roughness and material with subsurface scattering (subsurface parameter set to 0.5 and radii 1.0,0.2,0.1 in the blender PBR settings). These sets are reported in Table 3.3. This allows us to compare the performances of the PS methods when specific material features are changed, e.g., albedo, roughness, subsurface scattering. It is also possible to understand how the methods are robust to shape variations, locally by plotting error as a function of the normal/view angle, by globally comparing the results on the different objects of different complexity. One way to evaluate the object complexity is to evaluate the amount of shadow created on the different images that can be evaluated on the related rendering pass. For the three untextured objects, the total percentages of shadowed pixels in the 77 images are $0.20\%, 23.88\%, 16.68\%$, respectively. Figure 3.7 shows example renderings of the same object with the assigned 9 uniform materials, illuminated from the same light direction.

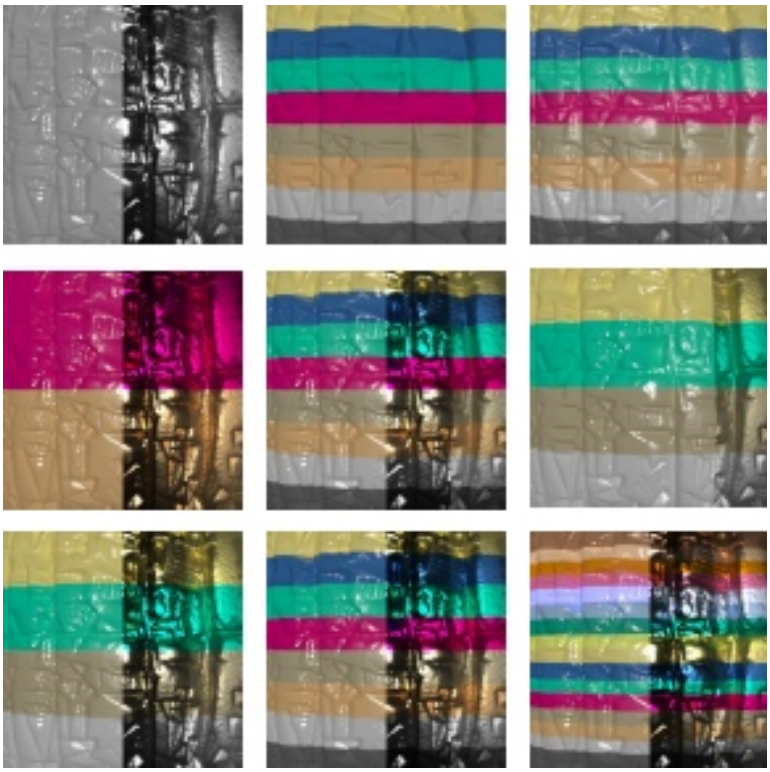


Figure 3.4: *Sample images of the cuneiform tablet with combinations of materials assigned*

Figure 3.5: *Rendered multi-light image collections (a) come with associated normal maps (b) that can be used for Photometric Stereo algorithms evaluation, shadow maps (c), that can be used for shadow detection testing and specularity maps (d)*

#	material	albedo	metal	spec	rough	subs
1	matte white (MW)	0.8	0	0	0	
2	white plastic smooth (PLA_WS)	0.8	0	1	0.4	
3	white plastic (PLA_W)	0.8	0	1	0.5	
4	white plastic rough (PLA_WR)	0.8	0	1	0.6	
5	gray plastic (PLA_G)	0.5	0	1	0.5	
6	dark plastic (PLA_D)	0.2	0	1	0.5	
7	metal smooth (MET_S)	0.8	1	0	0.5	
8	metal rough (MET_R)	0.8	1	0	0.7	
9	subsurface (SUB)	0.5	0	1	0.5	*

Table 3.3: *Parameters of the Cycles principled BSDF model used to create the 9 materials of the SynthPS, single material dataset.*

Figure 3.6: *Geometrical models used to create the SynthRTI and SynthPS datasets. Top left: untextured models used for the SynthRTI(SingleMaterial and MultiMaterial) renderings and SynthPS SingleMaterial renderings. Bottom right: textured models used for the SynthPS MultiMaterial renderings.*

The MultiMaterial dataset has been created with two different textured geometric models of CH items downloaded from SketchFab as well(`https://sketchfab.com`). All three have been digitized by G. Marchal and distributed under the Creative Commons 4.0 license. The first is the reconstruction of The lion of Goddess Ishtar, from

Nye Carlsberg Glyptotek. The second is an 11th century, sandstone, Lintel, Kapilapura from Angkor, Cambodia,11th century, from Musée du Cinquantenaire (Brussels, Belgium). For these models, we kept the original texture as the diffuse texture in the material settings and then used the Voronoi Texture node of Blender to define a spatial pattern used to map different roughness values for the material in different patches, ranging from 0.4 to 0.7 (see Figure 3.8). The SynthPS dataset can be accessed at the following URL:`https://github.com/Univr-RTI/SynthPS`.

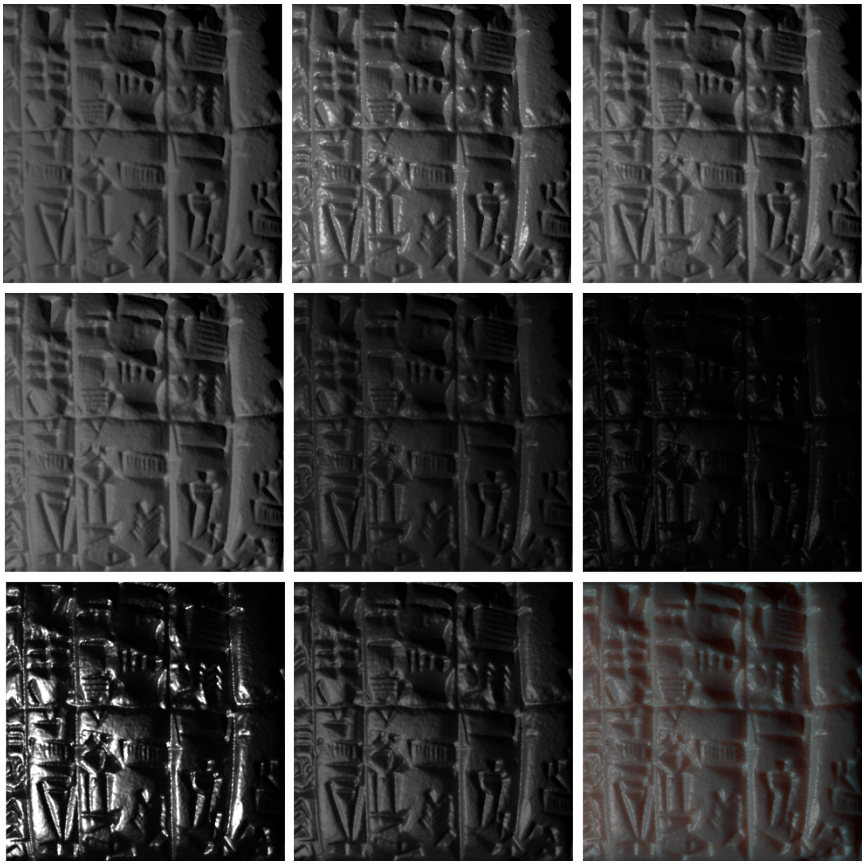


Figure 3.7: *Sample images of the same CH object with the assigned 9 materials and lit by the same illumination direction.*

3.1.4 Synthetic MLIC creation

To create the Multi-Light Image Collections(SynthRTI and SynthPS), we used a specific Blender script(plugin) allowing the automatic creation of light configurations from text files with light directions information and the automatic rendering of image sequences. Images are rendered with Blender Cycles engine, and all the solutions available in Blender for image quantization and tone mapping can be applied to simulate different acquisition

Figure 3.8: *Left: the Voronoi texture associated with the assigned roughness. Center, right: two renderings with different directional lights of the surface with the albedo texture and the regional roughnesses.*

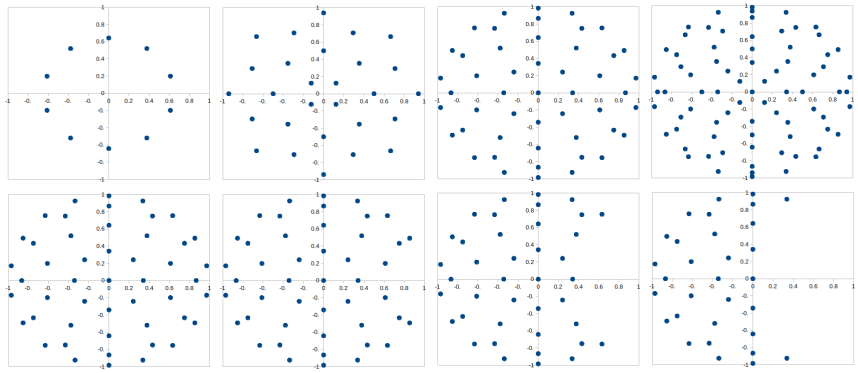


Figure 3.9: *Dome light configuration used for SynthPS benchmarking. Top: symmetric domes, from left, 10 light ring at 50 deg. elevation, 28, 49 and 77 directions domes. Bottom: asymmetric configuration created by removing lights from right in the 49 lights dome. From left: original, 46 lights, 40 lights, 31 lights represented in the l_x, l_y plane.*

and preprocessing methods.

For SynthRTI we simulated standard 8-bits preprocessed images as those generated by standard digital cameras, but corresponding raw images with 16 or 32 bits encoding could be generated as well. Whereas for SynthPS we simulated standard 16-bits preprocessed image. Figure 3.10 shows the plugin used to generate renderings.

The key features of the plugin are the following:

- Use of exactly directional lights, with sets of lights loaded from .lp files (use of point or spotlights could be applied to evaluate calibration issues).
- Loop over material parameters: this feature allows the generation of rendered objects of different materials within specified ranges of material parameters. This will be a fundamental requirement for the generation of training sets for material classification and segmentation
- Export of multiple images exploiting the different Cycles rendering passes: compos-

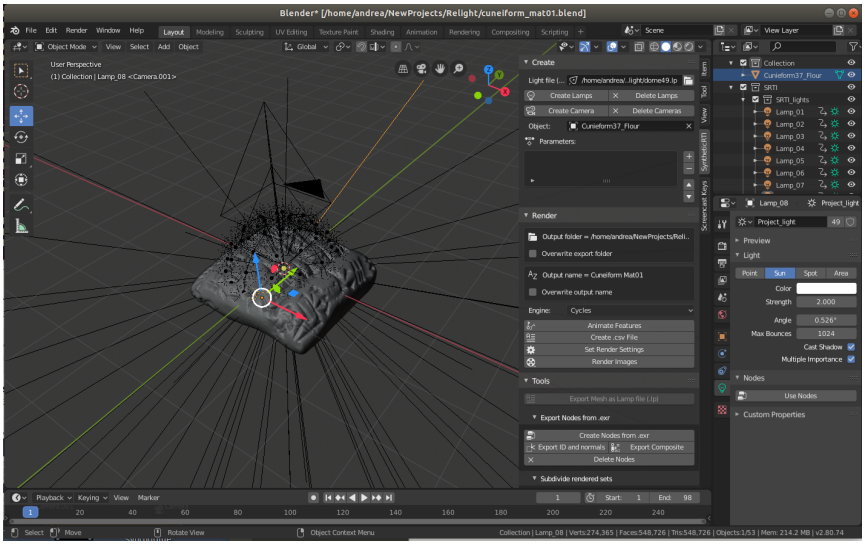


Figure 3.10: *Blender tool for the automatic creation of RTI dataset given light direction files.*

ite image, Specularity image, shadow image, material indexes. This will be useful to develop methods for shadows detection and/or robust fitting of a matte component.

The plugin is available at the website `https://github.com/giach68/SyntheticRTI`.

3.1.5 Photometric Stereo benchmarks

The most popular PS benchmark is currently the DiLiGenT dataset[132]. It is composed of captured images of real objects made of different materials, together with accurate metadata about light and camera calibration. The dataset has been used in a large number of papers, and with calibrated and uncalibrated PS methods. Sablatnig and Wimmer [10] used a dataset of ancient coins to evaluate the accuracy of a single PS method when the light sampling is varied. However, no details on ground truth estimation are provided and those data have not been publicly released. Xiong et al. [161] created a PS dataset with seven relatively-diffuse objects, captured with a CanonEOS 40D camera under directional lighting, and calibrated with chrome spheres [161]. Alldrin et al. [4] acquired high-dynamic-range images of two known objects in a dark room and calibrated the lights by employing reflective spheres.

While the use of real, calibrated images allows the evaluation of the methods on real materials and illumination, there are still few drawbacks. Light intensity and direction need to be calibrated, and it is not possible to avoid calibration errors. The camera model is not orthographic and lights are not directional but are approximately point lights. Shi et al. [132] provide both light positions and camera parameters, but often they are not used by many tested algorithms, which assume the orthographic camera model and directional lights. This may introduce a bias in the evaluation. Moreover, real lights are typically not uniform, and the intensity calibration can introduce errors as well. The objects and the

acquisition setups used to create the benchmarks are not exactly similar to those employed in the classical handheld RTI or light dome acquisition methods typically used in the Cultural Heritage domain. Furthermore, by using real acquisitions it is not possible to control and modify the local properties of the materials so it is hard to evaluate how the different PS methods behave with varying reflectance functions and spatial variations of them.

A multi-light images dataset with synthetic rendering (CyclesPS) has already been proposed, for example in [60]. However, it only features non-realistic objects with different materials assigned in superpixels, used to train (and test) local CNN-based PS algorithms.

This motivated us to create SynthPS, i.e., a specific dataset with synthetic, physically-based renderings of surfaces that can be considered typical examples of Cultural Heritage objects, which are made of different homogeneous and heterogeneous materials and have different geometrical complexity.

These features make the proposed dataset also suitable to design specific evaluation tests, testing the robustness of the different methods to a variety of specific factors (materials, material uniformity, depth variations, etc.) that are useful to choose the best approach for practical applications.

We rendered the objects with different configurations of uniform directional lights, by changing their number and their spatial configuration. We use this dataset to evaluate several recent PS approaches, both in terms of their accuracy and robustness against different factors(chapter 4). The resulting contribution consists of several guidelines to choose the proper image capture strategy and processing algorithm, together with a public benchmark that can be used by researchers to evaluate novel PS methods.

3.2 Summary

In this chapter, we proposed novel benchmark datasets that can be used to evaluate MLIC based methods, Reflectance Transformation Imaging, image relighting, photometric stereo and material classification/segmentation. The benchmark consist of synthetic, SynthRTI and SynthPS, simulated using a physically based rendering engine and real acquisitions, RealRTI.

SynthRTI is generated considering, mainly, RTI methods but can be used to evaluate image relighting and photometric stereo algorithms. And SynthPS is generated considering Photometric stereo algorithms for the applications of cultural heritage but can be used to evaluate image relighting, RTI and other MLIC based algorithms. SynthPS is composed of synthetic images simulating acquisitions of surfaces with different geometrical complexity, homogeneous and heterogeneous materials and assuming perfectly directional lights and orthographic view, thus avoiding inaccuracy in light or camera calibration. The image sets can be used to simulate light domes configurations with different densities and asymmetric layout simulating the effect of obstacles preventing the positioning of the lights from one side of the object.

We have used SythRTI and RealRTI to evaluate a neural network based RTI method NeuralRTI, [38] and SynthPS to evaluate Photometric Stereo algorithms to reconstruct surfaces of objects typically captured in multi-light acquisitions performed in the Cultural Heritage domain, described in chapter 4.

CHAPTER 4

Evaluation of Photometric stereo algorithms for Cultural Heritage application

As the capture of MLIC data is cheap and quite popular in the Cultural Heritage (CH) domain, Photometric Stereo(PS) could be a really useful tool to acquire high-resolution geometrical detail of the objects. However, it is not very clear which are the algorithmic choices that should be adopted to optimize the reconstruction quality. PS algorithms are typically evaluated on few benchmarks not particularly representative of typical CH applications, and they do not allow the evaluation of the variability of the methods' performances as a function of material types, camera and light calibration error, and image pre-processing(section 3.1.5).

In this chapter, we have used out SynthPS benchmark dataset(section 3.1.3) to test several PS methods, ranging from classical ones, to robust approaches, and those based on (deep) neural networks. The main reasons for this selection are related to the popularity of the chosen methods, the fact that they exhibit good results in existing benchmarks, and the public availability of their implementations.

4.1 Evaluation

4.1.1 Tested algorithms

The selected PS algorithms based on model fitting are standard Least Squares fitting of Lambertian model (LS) [154]; Trimmed Least Squares (Trimmed), i.e., LS fed with pruned

data, typically obtained by removing saturated values and a fixed percentage of high, possibly non-Lambertian, measures (in our tests 5%); Least Median of Squares(LMS) [36, 107]; Bayesian Regression(BR) [62]; Low-rank matrix completion and recovery (LMR) [155]; Sparse Regression [63].

Among deep learning algorithms, we either choose local or global methods; in the former ones, the normals are obtained on a per-pixel basis, while the latter train the network on the entire image domain. Furthermore, we also considered some uncalibrated methods (i.e., unknown input light directions). The tested deep learning based algorithms are: PS-FCN [18] (calibrated / global); CNN-PS [60] (calibrated / local); SDPS-Net [17] (uncalibrated / global). PS-FCN is a CNN based architecture that accepts a set of images as input, extracts features for each image, aggregates them via max-pooling, and finally infers the normal map. The method does not require a pre-defined set of light directions during training and testing, and it can handle multiple images and light directions in an order-agnostic manner. CNN-PS accepts as input an observation map, which is a 32x32 grid that encodes light intensity as a function of the light direction. A wide set of synthetic images is used in the training phase. Several different maps are generated in the prediction step, forcing a rotational invariance nature of the normal map estimation, which makes the method more robust. SDPS-net is a two-stage CNN based architecture. The first stage estimates the light directions, while the second stage uses the result of the first stage together with the captured images to compute the final normal map. Despite being uncalibrated, the method exhibits state-of-the-art performances on the DiLiGenT benchmark and performs well on challenging surfaces. However, it assumes surfaces with non-negligible relief and made of homogeneous material.

4.1.2 Results

The comparison of the methods is based on the statistics of angular error. For each pixel, the angular error is calculated as $\arccos(n_o^T n)$ in degrees, where n_0 and n are 'ground truth' and estimated normals respectively.

4.1.2.1 Uniform materials

For the three uniform material objects, the normal reconstruction errors for the typical 49-light dome configuration are reported in Table 4.1. It is possible to see that the accuracy of "global" neural methods is low for the nearly planar objects, while the local method (CNN-PS) provides good results like robust methods. CNN-based methods are far better than model-fitting techniques on normal-varying objects (bas-relief and tablet); although it explicitly needs shadows and relevant shading to solve for the light directions, the uncalibrated approach (SDPS) provides the best results.

Looking at the performances on different materials (Figure 4.1), it is possible to see that the performances of the best methods for each category are almost the same for matte and plastic materials, independently of roughness and are in some cases improved with smaller albedo values. Plots are done for the bas-relief results but are similar for the tablet. The performances of robust fitting methods are poor in the case of metallic surfaces. Neural methods are far better in this case, with SPDS-net (uncalibrated/global) generally providing the best results. Subsurface scattering is instead deteriorating the performances of all the methods in the same way.

	LS	Trim.	LMS	BR	LRM	SR	PSFCN	CNNPS	SDPS
MAT	0.35	0.56	0.83	0.28	0.26	**0.24**	1.58	0.26	17.91
PLA_WS	0.33	0.59	0.39	0.40	0.26	**0.24**	1.72	0.33	14.44
PLA_W	0.32	0.50	0.40	0.25	**0.23**	0.24	1.10	**0.23**	17.38
PLA_WR	0.32	0.40	0.41	0.37	0.40	**0.23**	1.83	0.25	23.91
PLA_G	0.37	0.77	0.40	0.73	0.71	0.29	0.67	**0.25**	42.08
PLA_D	0.59	1.69	0.56	2.52	2.58	0.50	3.11	**0.24**	33.60
MET_S	0.79	3.45	0.77	4.25	4.35	**0.68**	1.87	0.76	7.13
MET_R	0.54	0.67	0.62	1.62	1.69	0.46	2.24	**0.23**	42.27
SUB	0.42	0.74	0.57	0.43	0.45	0.35	1.34	**0.33**	23.03
Average	0.45	1.04	0.55	1.21	1.21	0.36	1.72	**0.32**	24.64

(a) nearly planar canvas

	LS	Trim.	LMS	BR	LRM	SR	PSFCN	CNNPS	SDPS
MAT	17.40	16.57	12.86	13.59	12.82	18.36	10.99	9.78	**9.35**
PLA_WS	16.75	16.90	13.34	14.10	13.44	18.54	11.55	10.68	**9.73**
PLA_W	17.01	16.83	13.29	14.09	13.44	18.47	11.36	10.36	**9.90**
PLA_WR	17.31	16.80	13.25	14.11	13.45	18.48	11.40	**10.17**	10.69
PLA_G	14.71	15.03	12.57	12.87	12.30	16.89	10.08	9.93	**7.00**
PLA_D	14.85	15.15	13.70	13.64	13.27	15.42	10.53	10.84	**9.95**
MET_S	18.23	19.52	22.07	18.60	18.71	18.14	12.12	13.76	**8.42**
MET_R	15.12	16.17	16.42	15.20	15.06	16.55	11.59	11.52	**10.21**
SUB	19.67	20.05	17.29	17.89	17.46	20.85	16.91	**14.77**	17.71
Average	16.78	17.00	14.98	14.90	14.44	17.97	11.84	11.31	**10.33**

(b) bas-relief

	LS	Trim.	LMS	BR	LRM	SR	PSFCN	CNNPS	SDPS
MAT	11.98	9.26	6.76	7.84	7.18	12.06	7.05	**6.61**	9.59
PLA_WS	10.67	9.55	7.18	8.11	7.58	11.84	7.26	7.14	**6.68**
PLA_W	10.98	9.52	7.19	8.14	7.61	11.77	7.05	**6.69**	7.45
PLA_WR	11.37	9.42	7.21	8.20	7.67	11.79	7.07	**6.47**	7.84
PLA_G	9.46	9.29	6.49	7.49	6.98	10.67	6.61	6.79	**4.29**
PLA_D	12.79	14.01	7.96	9.16	8.43	10.18	6.88	7.46	**4.15**
MET_S	16.83	18.59	14.03	13.43	12.77	12.41	7.65	9.15	**5.34**
MET_R	11.11	13.54	10.42	10.64	10.50	10.75	8.20	7.36	**5.19**
SUB	13.60	13.63	11.59	11.93	11.58	14.47	10.96	**9.63**	10.84
Average	12.09	11.87	8.76	9.44	8.92	11.77	7.64	7.48	**6.82**

(c) tablet

Table 4.1: *Mean angular error (deg.) for the normal reconstructions on the three objects of the SingleMaterial dataset with simulated 49 light dome and all the assigned materials. Bold fonts indicate the best results.*

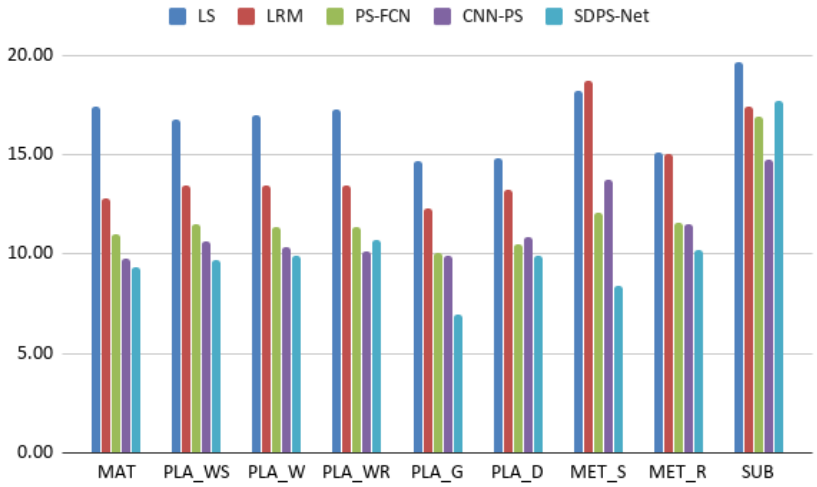


Figure 4.1: *Average errors for selected techniques: basic Lambertian (LS), a robust fitting (LRM), calibrated global network-based method (PS-FCN), calibrated local network-based method (CNN-PS) and uncalibrated network-based method (SDPS-Net) on the 49-lights normal recovery for the bas-relief models as a function of the different uniform materials.*

Let us now look at the effect of dome light density. Figure 4.2 shows the average angular errors on the bas-relief dataset with varying density. It is possible to see that the average errors are nearly the same by reducing the light density to 28 lights in a reasonably regular distribution, even if not all the methods are robust against the replacement of the dome configuration with the 50 degrees ring proposed by Sablatnig and Wimmer [10]. Neural methods are the best, but CNN-PS is not effective in the ring configuration (10 lights). This is due to the encoding of the pixel information as a 2D "observation map" that requires a reasonable density. The uncalibrated method performances are also decreased in the ring configuration as a larger number of lights are required for calibration. Neural methods are also more robust against asymmetry of the light direction configuration as shown in Figure 4.3.

It might be surprising to note that the uncalibrated method performs well despite the error in the light direction evaluation. However, it is possible to see that the addition of a randomly directed fixed-angle noise to the input light directions does not alter too much the accuracy of the methods as shown in Figure 4.4. The only method that seems to require accurate light directions is CNN-PS.

This fact also explains the success of the uncalibrated SPDS-Net method: as shown in Table 4.2 the average accuracy of the light direction estimated by its LCNet subnet estimating the light directions is not too accurate: fails on Object 1 and has a MAE close to 10 degrees on the depth varying surfaces, but the normal estimation subnet is robust against the error in the input light direction when it is lower than 10 degrees.

The average errors on the entire images, however, do not show how the performances

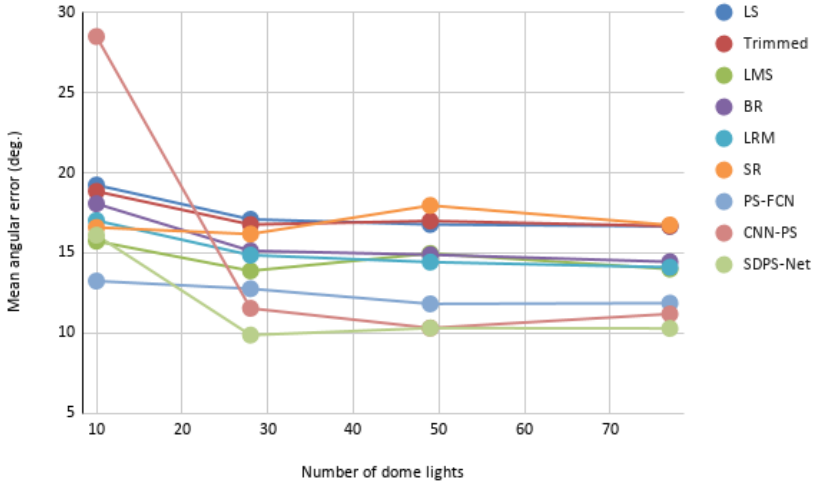


Figure 4.2: *Average mean angular errors (on all the different material) for the bas-relief normal estimation vs number of lights in radially symmetric configurations.*

of the methods change in critical regions, for example where the normals create large angles with the view directions and where nonlocal effects like shadows are relevant.

To analyze the performance of the methods in challenging regions we can look at error maps, like those shown in Figure 4.5, and plot methods accuracy against local properties, e.g., the z-component of the ground truth normals, or the percentage of shadowed light directions, which can be obtained from the synthetic rendering output (see figure 3.5). We report in Figure 4.5 the error maps for the bas-relief 49-lights dome test, material 3 (white plastic).

It is possible to see that, as expected, the error is higher where the surface is not flat and the effect of shadows and inter-reflections is higher; this is more evident for Lambertian model-fitting methods. This fact can be quantitatively shown by plotting the errors as a function of local surface properties. Figure 4.6 shows the average pixel errors on the 49-lights capture of the bas-relief model (average on all the materials) for pixels as a function of the angle between the actual surface normal and the view direction. It is possible to see that, while for surfaces nearly perpendicular to the view direction all the methods are similarly accurate, the errors grow with the angle at different ratios, with neural methods far better at large angles. Figure 4.7 shows the average pixel errors on the 49-lights capture of the bas-relief model (average on all the materials) as a function of the percentage of locally shadowed lights (known from the rendering step). The error is similar for all the methods and low when shadowed directions are less than 20%, then there is a rapid growth with neural methods far better. CNN-PS seems more robust when the percentage exceeds 70%. SDPS-Net, despite not using input light directions, provides the best results, even if the largest errors seem more relevant than those of other neural methods in Figure 4.5. As reflected in Figure 4.7, these points correspond to the ones with a large percentage of shadowed pixels.

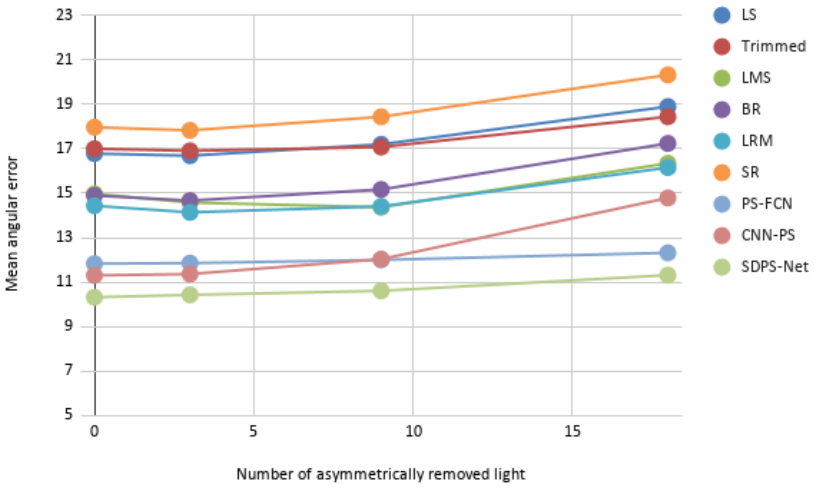


Figure 4.3: *Average errors on bas-relief normals estimation vs number of lights removed from a side of the 49-lights dome.*

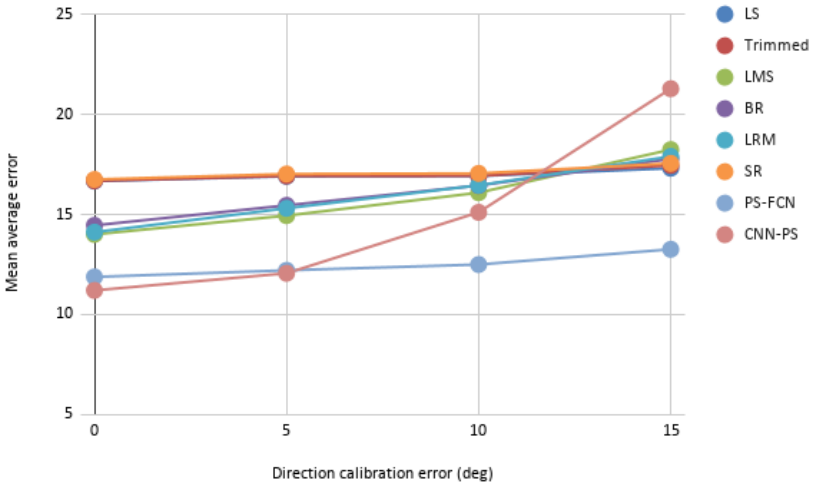


Figure 4.4: *Average errors on bas-relief normals estimation vs simulated error in light direction calibration.*

	Object1	Object2	Object3
MAT	57.60	6.35	9.22
PLA_WS	61.71	8.33	8.39
PLA_W	58.44	8.06	8.51
PLA_WR	57.97	7.31	8.77
PLA_G	60.47	7.68	8.11
PLA_D	65.98	9.14	8.97
MET_S	59.76	12.03	10.20
MET_R	60.20	9.51	8.67
SUB	60.14	16.79	10.56
Average	60.25	9.47	9.04

Table 4.2: *Light direction estimation errors obtained on the 49-lights image collections with the self-calibration module of SDPS-Net on the three uniform surfaces with different materials assigned*

	LS	**Trim.**	**LMS**	**BR**	**LRM**	**SR**	**PSFCN**	**CNNPS**	**SDPS**
Object1	5.52	5.21	4.38	4.81	4.67	5.19	6.99	**4.04**	12.10
Object2	12.23	11.80	9.28	9.73	9.17	12.62	9.74	**7.92**	19.87
Average	8.88	8.51	6.83	7.27	6.92	8.92	8.37	**5.98**	15.99

Table 4.3: *Average of the MAE obtained by the different methods on the two multimaterial objects. Bold fonts indicate the best results.*

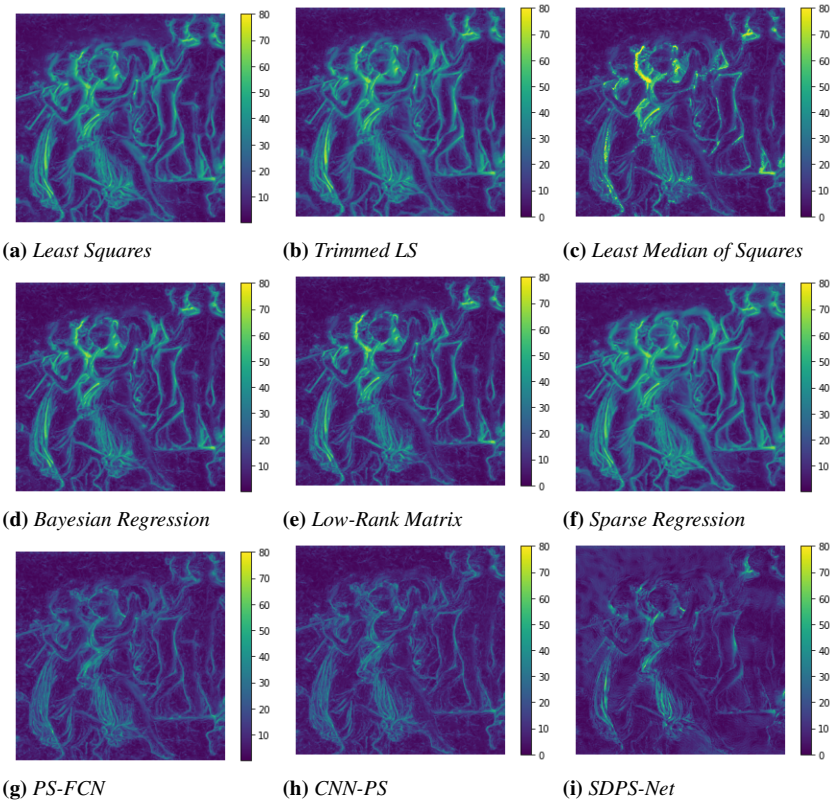
4.1.2.2 Non-uniform materials

When the materials are not uniform, the ranking of the methods is completely different. As expected, because the hypothesis of uniform material is exploited in the direction calibration network, SPDS-Net fails, but also PS-FCN is no longer performing well. CNN-PS is clearly the best one, probably due to its local nature, even if robust fitting methods are not too far, as shown in Table 4.3.

It is interesting to see in Figure 4.8 that the accuracy of CNN-PS is practically constant when the number of the directional dome light is reduced from 77 to 28 lights uniformly distributed, but this method completely fails if the dome is replaced with a single-elevation ring. This is due to the impossibility of estimating a dense observation map exploiting rotational symmetry. Least Median of Squared robust fitting, on the other hand, seems sufficiently accurate also in this case, even if not as in the uniform light dome configuration.

Another limit of the CNN-PS method, at least with the training provided by the authors, is related to the symmetry of the input light set. In onsite handheld acquisitions typical of the Cultural Heritage domain it is rather usual that lights cannot be placed on a side of the surface of interest. However, CNN-PS is sensitive to the asymmetrical removal of lights as shown in Figure 4.9

Looking at the error maps for the Lion 49-light capture (Figure 4.10) we can see that the error of SDPS-Net is strongly influenced by the background roughness (patches are visible in the error map). CNN-PS is, clearly, the most effective technique, even if robust fitting methods are effective as well.



(a) *Least Squares* **(b)** *Trimmed LS* **(c)** *Least Median of Squares*

(d) *Bayesian Regression* **(e)** *Low-Rank Matrix* **(f)** *Sparse Regression*

(g) *PS-FCN* **(h)** *CNN-PS* **(i)** *SDPS-Net*

Figure 4.5: *Error maps encoding local errors in normals for the white plastic bas-relief captured with the simulated 49-lights dome and reconstructed with different PS algorithms.*

The behavior of the methods is different with respect to the surface normal or the percentage of shadowed directions. Figure 4.12 shows that many robust fitting methods (but not LMS) present a strong growth of the error with the angle between normal and view direction, while the error of SDPS-Net is large but mostly unaffected by normal direction.

If we look at the effect of shadowed directions, CNN-PS is constantly the best method independently of the number of shadowed directions. LMS starts to fail when the number of shadow outliers exceeds 40%, while LRM behaves well and is quite close to CNN-PS at large values (see Figure 4.11).

4.1.2.3 Effect of different image encoding

In our synthetic image encoding, we did not apply gamma correction and we recorded a linear signal, encoded with 16 bits per channel, simulating raw images captured with a sensor with a high dynamic range. However, multi-light image captures often are performed with low-cost hardware and may be encoded as 8 bits images, possibly with gamma cor-

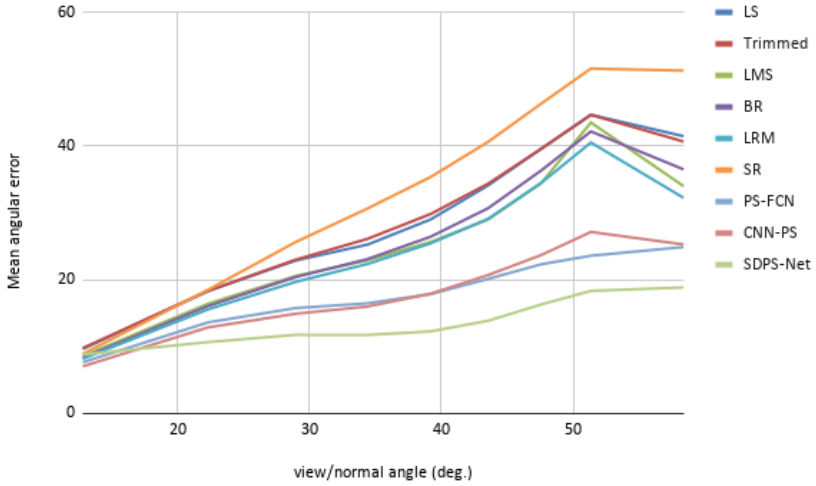


Figure 4.6: *Average MAE (all materials) obtained with all the tested methods on the pixels of the bas-relief 49-lights image collections vs angle between normal and view direction.*

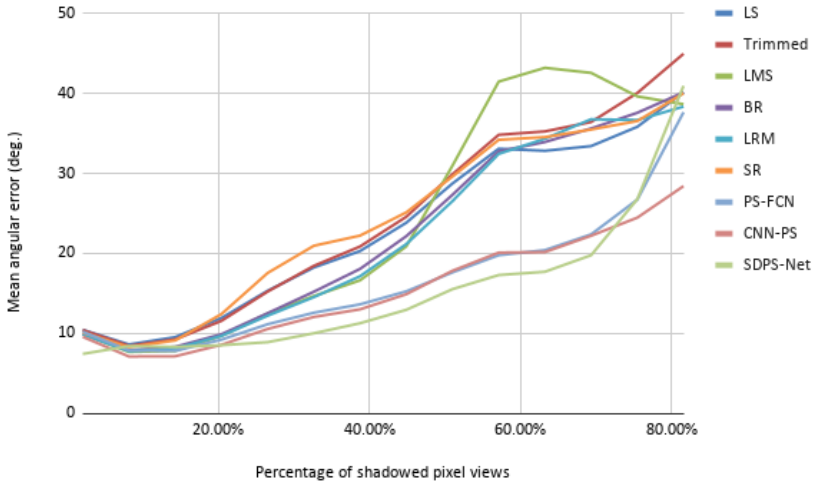


Figure 4.7: *Average MAE (all materials) obtained with all the tested methods on the pixels of the bas-relief 49-lights image collections vs percentage of shadowed directions.*

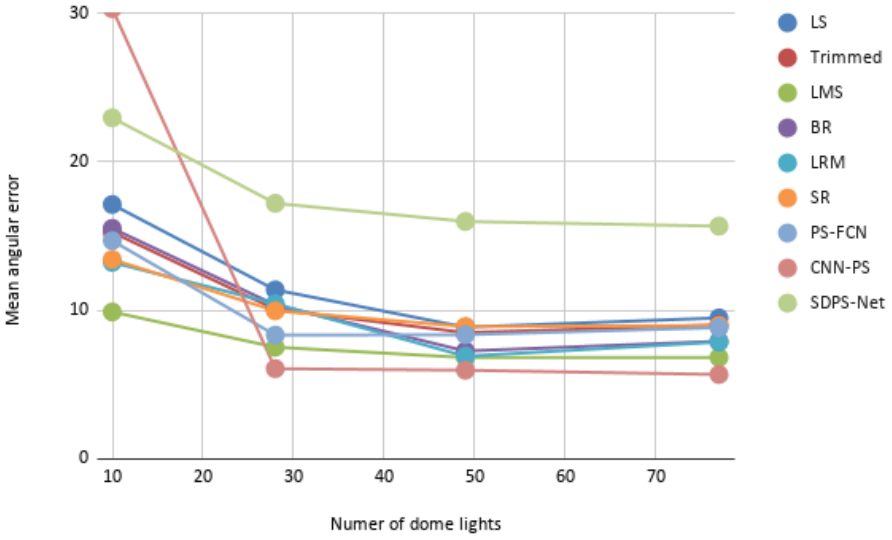


Figure 4.8: *Average MAE on the two multimaterial objects vs number of lights in a symmetric dome configuration.*

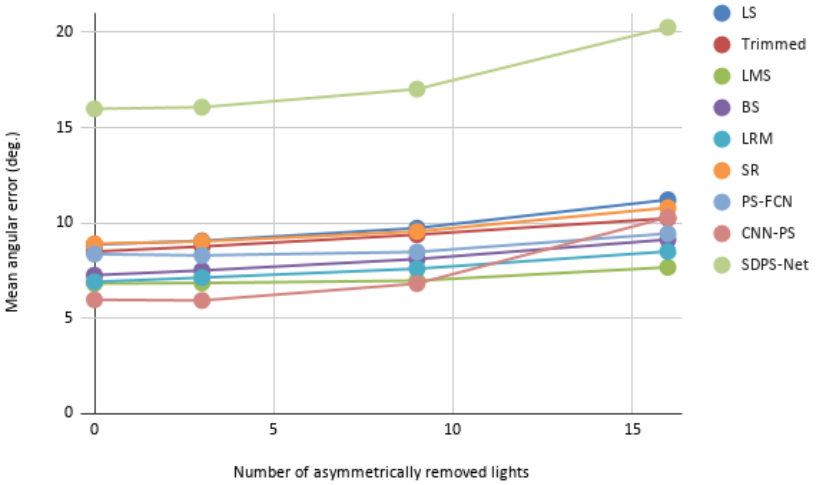
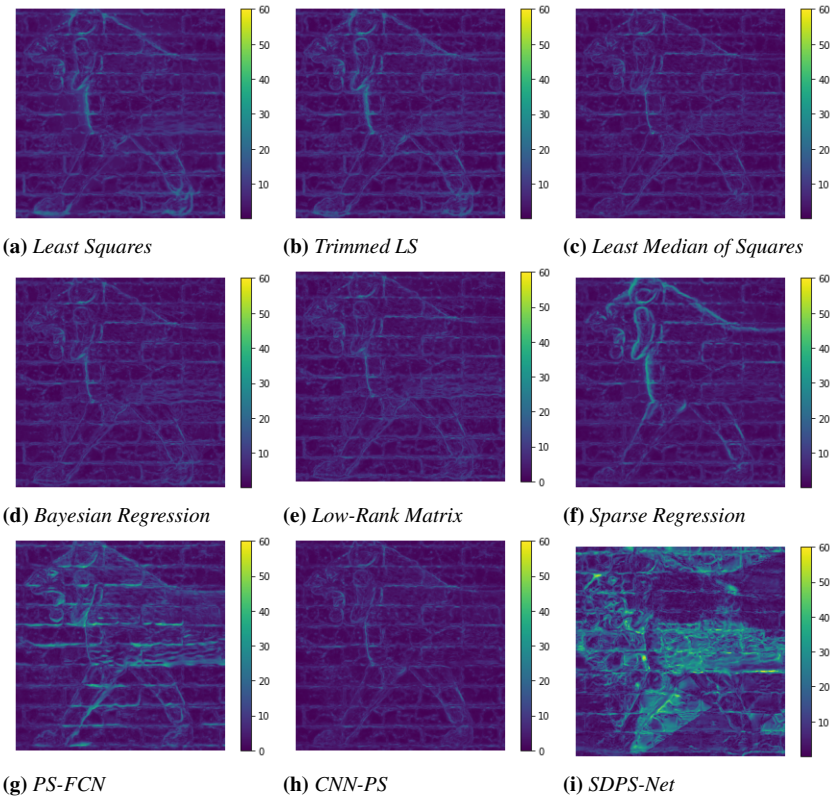


Figure 4.9: *Average MAE on the two multimaterial objects vs number of lights removed from a side of the simulated 49-lights dome.*

rection or unknown nonlinear mapping. We performed some tests also to verify the effects of this on the accuracy of the reconstructed normals performed with PS algorithms. Figure 4.13 shows the results obtained with the tested methods on the original 16 bits linear



(a) *Least Squares* **(b)** *Trimmed LS* **(c)** *Least Median of Squares*

(d) *Bayesian Regression* **(e)** *Low-Rank Matrix* **(f)** *Sparse Regression*

(g) *PS-FCN* **(h)** *CNN-PS* **(i)** *SDPS-Net*

Figure 4.10: *Error maps encoding local errors in normals for a multimaterial object captured with the simulated 49-lights dome.*

images (blue bars), on 8 bits linear images (red bars), 8 bits nonlinear images ($\gamma = 2.2$) remapped linearly before normal estimation, and 8 bits nonlinear images not corrected (purple bars). The results show that for simulated acquisitions the 8-bit discretization does not affect the results. With real acquisitions and non-ideal sensors, results can be different, but in this case, there is no need for keeping the full dynamic range. Furthermore, it is interesting to note that the best method (CNN-PS) is rather robust with respect to the lack of correct linearization of the input illumination.

4.2 Summary

In this chapter, exploiting the features of SynthPS we performed a set of tests on state of the art PS techniques, that can be used to derive suggestions or even guidelines for the practical reconstruction of normal maps and surfaces from MLIC data, for cultural heritage application. Here are the main outcomes of our tests:

- Neural methods are promising, providing the best results in most of the compar-

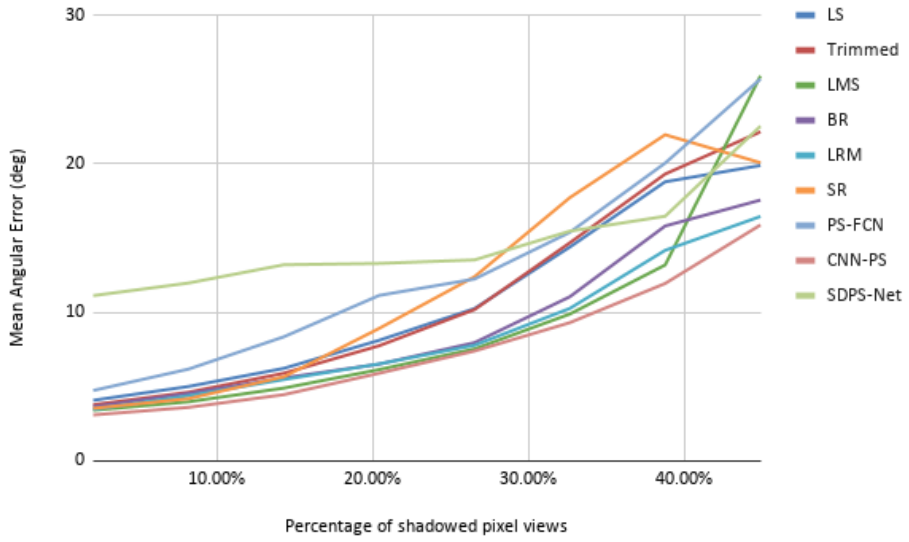


Figure 4.11: *Average MAE obtained on the pixels of the Lion 49-lights capture by the different methods vs percentage of shadowed directions.*

isons. However, due to intrinsic limits or training on limited surface types, no single method can cope with all the possible acquisition settings.

- SDPS-Net is quite effective for uniform materials and does not require light calibration. However, as clearly stated by the authors, it does not work on nearly-planar objects and multimaterial surfaces.
- PS-FCN even if calibrated is not accurate on nearly-planar objects and does not perform well on multimaterial objects. This may be due to the use of global information
- The methods based on robust Lambertian fitting works sufficiently well on matte and plastic materials, but less on metallic ones.
- CNN-PS seems the most reasonable choice in general, as it works well on nearly flat surfaces and it is the more effective in the case of non-uniform materials. However, it is more sensitive than the other techniques to the accuracy of light direction calibration and the asymmetry of the light layout. This fact is an intrinsic limitation of the method using "observation maps" as input but can be reduced using other learning based method to densify sparse input maps.
- Performances of the training-based methods depend on the training data used. Good performances of CNN-PS may be biased by the fact that it is trained with synthetic data rendered with Cycles as SynthPS. Performances of these methods can however be improved on different sets of data by enlarging the training set including different kinds of data.

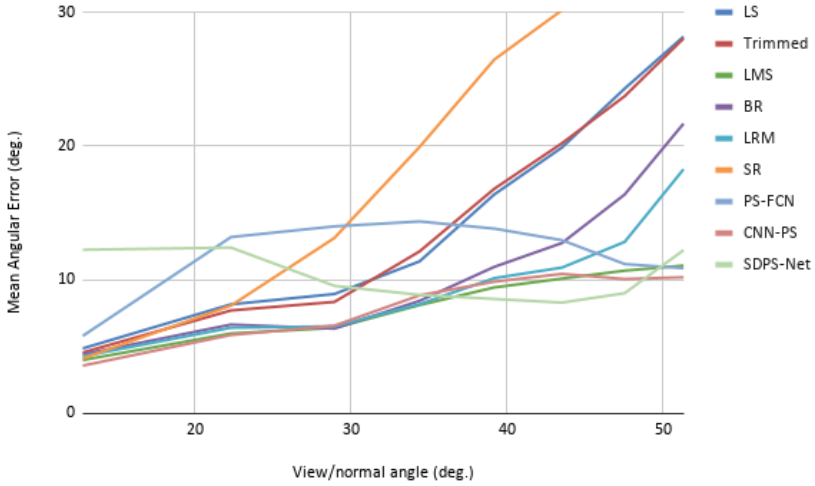


Figure 4.12: *Average MAE obtained on the pixels of the Lion 49-lights capture by the different methods vs angle between normal and view direction.*

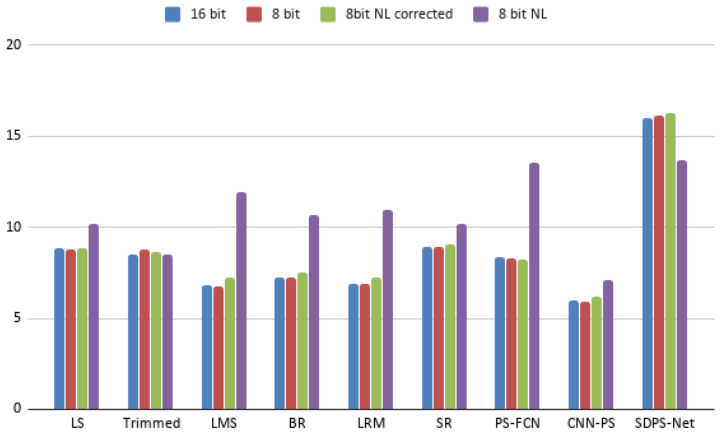


Figure 4.13: *Average MAE on the two multimaterial objects obtained with the tested methods on differently encoded images: 16 bits linear, 8 bits linear, 8 bits with standard gamma correction and inverse correction before fitting, 8 bit with gamma correction without linearization.*

- On well exposed images it seems to be not crucial to keep 12-16 bits depth in input images as there is no apparent decrease in performance with a simple linear mapping onto 8 bits.

CHAPTER 5

NeuralRTI:Neural Transformation Imaging

Reflectance Transformation Imaging(RTI) is a computational photography technique widely used in the cultural heritage and material science domains to characterize relieved surfaces. It consists of capturing multiple images from a fixed viewpoint with varying lights. Handling the potentially huge amount of information stored in an RTI acquisition, that consists of typically 50-100 RGB values per pixel, allowing data exchange, interactive visualization, and material analysis is not easy. The solution used in practical applications consists of creating "relightable images" by approximating the pixel information with a function of the light direction, encoded with a small number of parameters. This encoding allows the estimation of images relighted from novel, arbitrary lights, with a quality that, however, is not always satisfactory.

Despite the remarkable results obtained with neural network based methods on other tasks related to the processing of multi-light image collections, (e.g. Photometric Stereo [17, 165]) no practical solutions based on them have been proposed to replace classical RTI(PTM, HSH, DMD and PCA/RBF) methods.

In this chapter, we present NeuralRTI, a novel framework for pixel-based encoding and relighting of RTI data. Using a simple autoencoder architecture, we show that it is possible to obtain a highly compressed representation that better preserves the original information and provides increased quality of virtual images relighted from novel directions, especially in the case of challenging glossy materials. The experiments performed on SynthRTI and RealRTI datasets(chapter 3) demonstrate the advantages of the proposed relightable image encoding. The context of our work has presented in chapter 2, section(2.3).

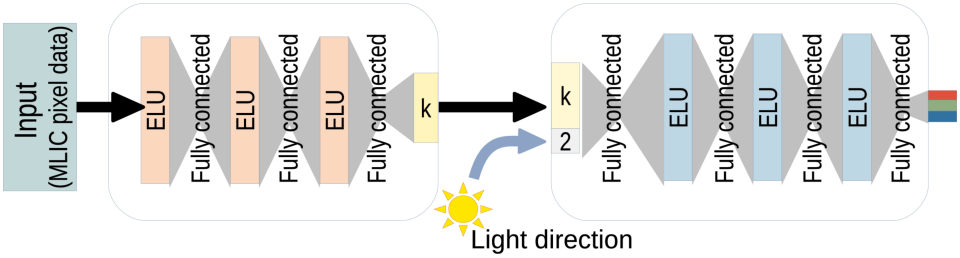


Figure 5.1: *Our neural RTI builds on an asymmetric encoder-decoder architecture. The encoder receives a per-pixel RTI data as input, i.e. a set of pixel values associated with a different lighting direction. These measurements are then passed through a sequence of fully connected layers until a fully connected layer finally produces a k-dimensional code. The decoder concatenates the code vector with the light direction and passes them through a sequence of fully connected layers with componentwise nonlinearities. The last layer outputs a single RGB color.*

5.1 Neural Reflectance Transformation Imaging

The basic idea of Neural RTI is to use the data captured in a multi-light image collection to train a fully-connected asymmetric autoencoder mapping the original per pixel information into a low dimensional vector, and a decoder able to reconstruct pixel values from the pixel encoding and a novel light direction.

Autoencoders [9] are unsupervised artificial neural networks that learn how to compress the input into a lower-dimensional code, called the latent-space representation, and then reconstruct an output from this representation that is as close as possible to the original input. In the recent past, autoencoders have been widely used for image compression [56, 166, 168] and denoising [74, 158], as they can learn non-linear transformations, unlike the classical PCA, thanks to non-linear activation functions and multiple layers. We specialized this kind of architecture to create an RTI encoding that is compact like currently used RTI files and can be used as well for interactive relighting, provided better quality and fidelity to the original data.

Our encoder/decoder network, represented in Figure 5.1 is trained end-to-end with the pixel data of each multi-light image collection to be encoded and the corresponding light directions as input. The training procedure minimizes the mean squared loss between predicted and ground truth pixel values on the set of given light directions.

The rationale of the network design is that we can then store the compact pixel code together with the coefficients of the decoder network (that are unique for all the pixel locations) as a relightable image. Given a light direction included in the original set, we can reconstruct the original images (compression), while using generic light directions we can generate relighted images. We assume that, if the sampling of the original data is reasonably good as in real-world RTI acquisitions, the network can provide better relighting results compared to traditional methods, as it can learn non-linear light behaviors and not only the coefficients but also the relighting function is adapted to the input data (shape and materials).

The encoder network (Figure 5.1, left) includes 4 layers, each consists of one activation

layer. The first three layers contain 3N units, where N is the number of input lights, and the last layer is set to the desired size k of the compressed pixel encoding (9 in our tests). Each layer is equipped with the Exponential Linear Unit (ELU) activation function. Since the encoder network is used only during the training, there are no particular constraints on the architecture or its size. However, from empirical analysis, we found that 9 coefficients seem sufficient to provide good relight quality and the size of intermediate layers of encoders and decoders equal to the input size provides good reconstruction results limiting overfitting.

The decoder network (Figure 5.1, right) consists of 3 hidden layers consists of 3N units each. The input is the concatenation of the pixel encoding and the 2D vector with the light direction. The output is the predicted single RGB pixel value, illuminated from the given light direction.

We used 90% of the total RTI data pixel as training data and a 10% sampled uniformly across pixel locations and light directions as the validation set. Training is performed with the Adam optimization algorithm [65] with a batch size of 64 examples, a learning rate of 0.01, a Gradient decay factor of 0.9 and a squared gradient decay factor of 0.99. The network is implemented using the Keras open-source deep learning library. After the training, the codes corresponding to pixel locations are converted to byte size with the offset+scale mapping and stored as image byte planes, and the coefficients of the decoder $(3N^2+5N)$ are stored as header information to be used as an input to the specific relighting tool and used at runtime to relight images give novel arbitrary light directions.

In our tests, we trained the RTI encodings on a GeForce RTX 2080 Ti machine with a single GPU, and training takes approximately one hour whereas the time required to create a 320×320 relighted image is approximately 0.007 seconds.

5.2 Results and evaluation

On the novel SynthRTI (section 3.1.2) and RealRTI (section 3.1.1) datasets, we performed several tests to evaluate the advantages of the proposed neural encoding and relighting approach with respect to standard RTI encoding.

5.2.1 Evaluation methodology

We followed two different protocols for the evaluation of the relighting quality on SynthRTI and RealRTI. On the synthetic data, we created the relightable images with the different methods using the *Dome* subsets and tested the similarity of the images of the *Test* subset with the images relighted with the corresponding light directions. The similarity has been measured with the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index (SSIM).

On the RealRTI, as test data are missing, we followed a leave-one-out validation protocol: for each of the image collections, we selected 5 test images with different light elevation. Each of them has then, in turn, removed from the collection used to create relightable images and we evaluated the similarity of the removed images with the images relighted from the corresponding light direction. We finally averaged the similarity scores of the 5 tests.

5.2.2 Compared encodings

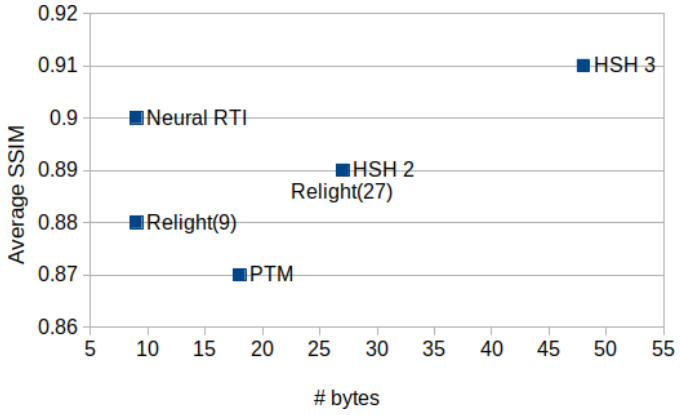
We compared the novel neural encoding with the RGB RTI encodings obtained with second-order Polynomial Texture Maps (PTM), second and third-order Hemispherical Harmonics RTI (HSH) and PCA-compressed RBF interpolation [116]. In PTM encoding, 6 fitting coefficients per color channel are obtained as floating-point numbers and compressed to byte size mapping the range between global max and min values into the 8 bits range. The total encoding size is therefore 18 bytes per pixel. In the second-order HSH, we similarly have an encoding with 9 coefficients per color channel and a total of 27 bytes per pixel. In the third-order HSH, we have 16 coefficients per color channel and a total of 48 bytes per pixel. We implemented the encoding and relighting methods using Matlab. Note that we could further compress the PTM/HSH encoding by using the LRGB version, using a single chromaticity for all the light direction. This choice, however, makes it not possible to recover the correct tint of metallic reflections. In the PCA-RBF encoding, it is possible to choose the number of PCA components used for the input data projection. We decided to test the results obtained with 9 and 27 components. We used the original "Relight" code provided by the authors of [116] to create the encoding and relight images. In the Neural encoding, as in PTM/HSH, we quantize coefficients to 8 bits integers with the min/max mapping so that they can be stored into image planes. This will allow easy integration of the novel format into existing viewers.

5.2.3 SynthRTI

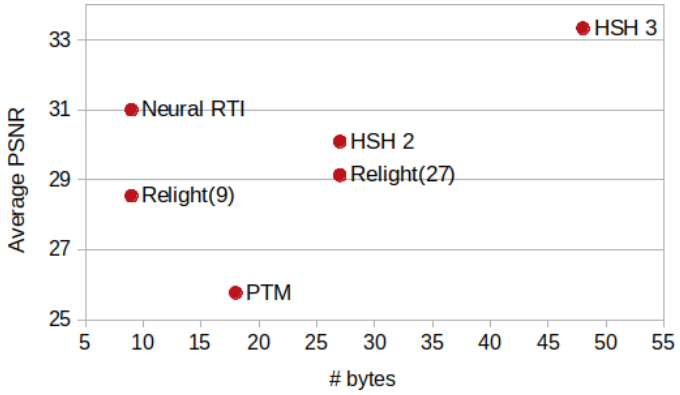
If we look at the SSIM and PSNR values obtained on average on the three shapes with the single materials, we see that the neural relightable images are quite good, as the average scores estimated on all the dataset (8 materials times 3 shapes) are quite high despite the low number of parameters used by the coding (9). Figure 5.2 shows a plot representing the scores as a function of the number of encoding bytes and NeuralRTI stays on the top left of the plot. Scores of 3rd order HSH are higher, but the method requires 48 bytes for the encoding and, as we will see, it creates more artifacts on challenging materials. Other methods to compress the relightable image encoding are not as effective as NeuralRTI. This is demonstrated by the poorer results of 2nd order HSH encoding, which can be considered a compressed version of the 3rd order one, and by the poorer results of the PCA-based compression of the full original information featuring 147 bytes per pixel (Relight).

Figure 5.3 shows the SSIM values for the different materials obtained by low-dimensional encoding methods as PCA/RBF (9 coefficients), Neural (9) and HSH 2nd order (27). Average SSIM values are close for HSH and Neural, but NeuralRTI performs better on challenging materials simulating dark plastic behavior and polished metals (5 and 6), despite the three times more compact code. Differences are statistically significant ($p < 0.01$ in a T-Test). HSH seems to handle significantly better the material with subsurface scattering effect on, probably thanks to color separation. Algorithms ranking does not depend on the geometric complexity of the model. Table 5.1 shows average PSNR values for the different models.

A visual comparison clarifies much better the advantages of NeuralRTI. Figure 5.4 shows relighting performed with 2nd order HSH (a), 3rd order HSH (b), PCA/RBF with 9 or 27 coefficients (c,d), NeuralRTI (e) on the canvas object with assigned white "plastic"



(a) *average SSIM*



(b) *average PSNR*

Figure 5.2: *Plots representing average SSIM and PSNR scores of the relighted images obtained with different methods from the test light directions of the SingleMaterial set. Scores are plotted versus the number of bytes used for the compressed encoding. Best methods should stay on top left.*

(material 2) behavior. (e) shows the ground truth image corresponding to the input direction. Only the Neural relighting (e) can reproduce the specular highlight with reasonable accuracy.

Figure 5.5 shows relighting results obtained with different methods on the bas relief shape with a metallic material (6) assigned, compared with the corresponding ground truth test image (f). It is possible to see that the highlights and shadows provided by the novel technique are the most similar to the real ones. HSH and PCA/RBF encodings with a limited number of parameters (a) and (c) appear matte. Adding more coefficients (b,d), the quality is better, but the contrast and the quality of the detail are not as good as in the result obtained with Neural RTI (e), despite the heavier encoding. Like the other methods, NeuralRTI fails in reproducing correctly all the cast shadows (see for example the one on the right of the woman on the top right). This is expected, as the method is local, but the

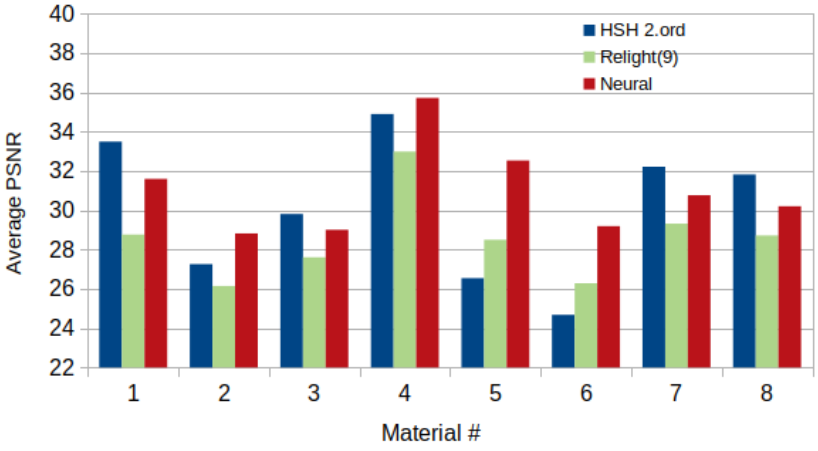


Figure 5.3: *Bar chart showing average SSIM of selected methods on the different materials. Neural relighting and 2nd order HSH perform similarly on rough materials, while neural relight is significantly better on glossy materials (5 and 6, see Table 3.1)*

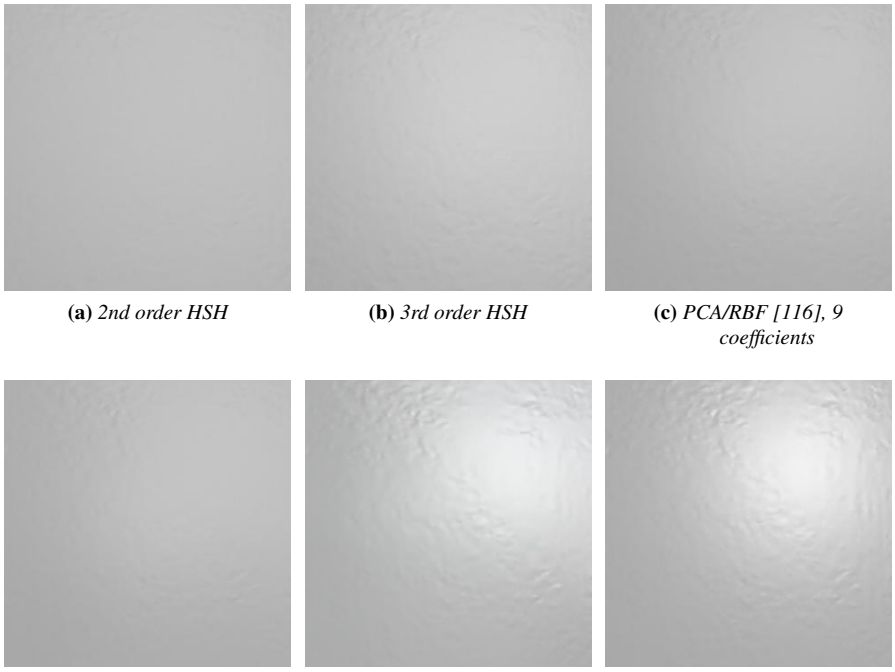
	PTM(18)	HSH(27)	HSH(48)	[116](9)	[116](27)	Neur.(9)
canvas	29.03	35.42	41.24	34.20	34.26	37.50
tablet	23.79	27.63	29.92	25.87	26.80	28.06
bas relief	24.47	27.22	28.82	25.55	26.31	27.43
avg	25.77	30.09	**33.33**	28.54	29.13	31.00

Table 5.1: *Average PSNR of the methods on the different shapes (all SingleMaterial sets). Parentheses include the number of coding bytes.*

material appearance is realistic on this challenging material.

Plotting the average relighting quality scores obtained on the MultiMaterial subset (Figure 5.6), we see that NeuralRTI compares favorably with the other methods, as, despite the compact encoding, it demonstrates the highest relighting quality. This is because many of the multi-material combinations include metallic and glossy materials, that are not rendered properly by the other methods. This fact appears clear looking at the bar chart in Figure 5.7. The NeuralRTI relighting has significantly higher SSIM values ($p < 0.01$ in T-tests) for all the material combinations except 2,3 and 4. 2 is fully matte, 3 is all simulating rough plastic, 4 is mostly composed of rough materials. Average PSNR values for the different geometric models are reported in Table 5.2.

A visual analysis of the results demonstrates here the advantages of NeuralRTI as well. Figure 5.9 shows relighted images obtained from the encodings made on the dome MLIC of material mix 5. Here we have two tints and, on the left, plastic behavior with 4 levels of roughness and, on the right, metallic behavior with 4 levels of roughness (increasing from left to right). 2nd order HSH fails in reproducing the correct highlights (a), compared with the ground truth ones (f). 3rd order HSH better represents specularity (b), but still with artifacts, and requires 48 bytes per pixel. PCA/RBF-relighted image also appears



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

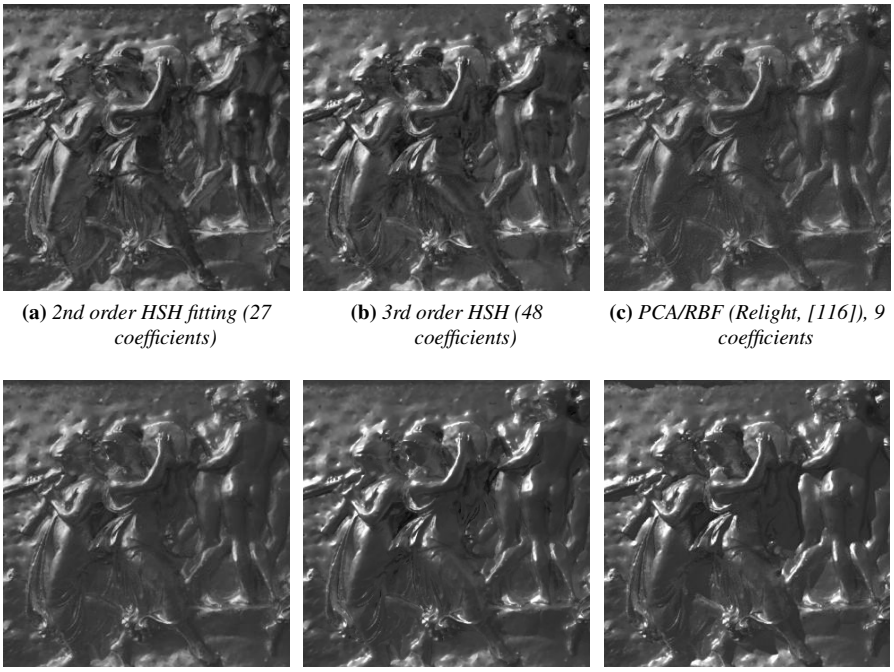
Figure 5.4: *Comparison of relighted images from simulated RTI data of a painted canvas, "plastic" material, light elevation 80 degrees.*

matte with 9 bytes encoding (c) and is only slightly improved using 27 coefficients (d). NeuralRTI provides a good relighting of all the image regions, creating an image that is quite similar to the reference one in the test set. The directional light, in this example, comes here from an elevation of 60 degrees. In general, the ranking of the relighting quality is not changed by the input light elevation. Quality scores are all higher for higher elevation values.

5.2.4 RealRTI

Tests on real images confirm the evidence coming from those performed on the synthetic ones (and this also shows that the rendered materials have reasonably realistic behaviors). Looking at the average SSIM and PSNR plots obtained on the whole RealRTI dataset, it is possible to see that Neural RTI provides the most accurate relighting even using only 9 coefficients. The difference in the average score is mainly due to the different quality of the relighting of shiny metallic objects. The bar chart in Figure 5.8 reveals that significant differences are found in items 9 and 10, which are metallic coins.

This fact can be seen by visually comparing relighted images with different methods. Figure 5.11 shows an image relighted with the leave-one-out procedure from a metallic coin acquisition. Relighting obtained with PCA/RBF with 9 coefficients shows a wrong



(a) *2nd order HSH fitting (27 coefficients)*

(b) *3rd order HSH (48 coefficients)*

(c) *PCA/RBF (Relight, [116]), 9 coefficients*

(d) *PCA/RBF (Relight, [116]), 27 coefficients*

(e) *NeuralRTI (9 coefficients)*

(f) *ground truth image*

Figure 5.5: *Comparison of relighted images from simulated RTI data of a shiny (metal smooth) bas-relief.*

tint in the central part, limited highlights, and halos, the use of 27 coefficients removes the tint issue. Image relighted with HSH shows a correct tint thanks to RGB decoupling, but relevant halos and missing highlights, especially in the 2nd order version. The NeuralRTI result is the only one with realistic highlights. Furthermore, it is possible to see that the proposed technique also avoids the typical artifact arising in the shadowed regions of RTI relighting, which appear as a blending of the shadows of the input images when obtained using PTM, HSH, and RBF. A detailed quantitative and qualitative results, on both SythRTI and RealRTI, can be found in the appendix A.

5.2.5 Comparisons with other network architectures

As we pointed out in Section 2.3, there are no available frameworks for the neural network based compressed encoding of relightable images, but there are, actually, methods for multi-light data relighting based on different network architectures, like those described [120, 162]. We tested the use of these architectures, adapted for our scope and compared the relighting accuracy with the one provided by our method but did not include the results in the previous sections as it would not be fair to compare encodings not specifically designed and used for this kind of sampling. We show our results here just to confirm that our NeuralRTI encoding seems particularly suitable for the task also considering different

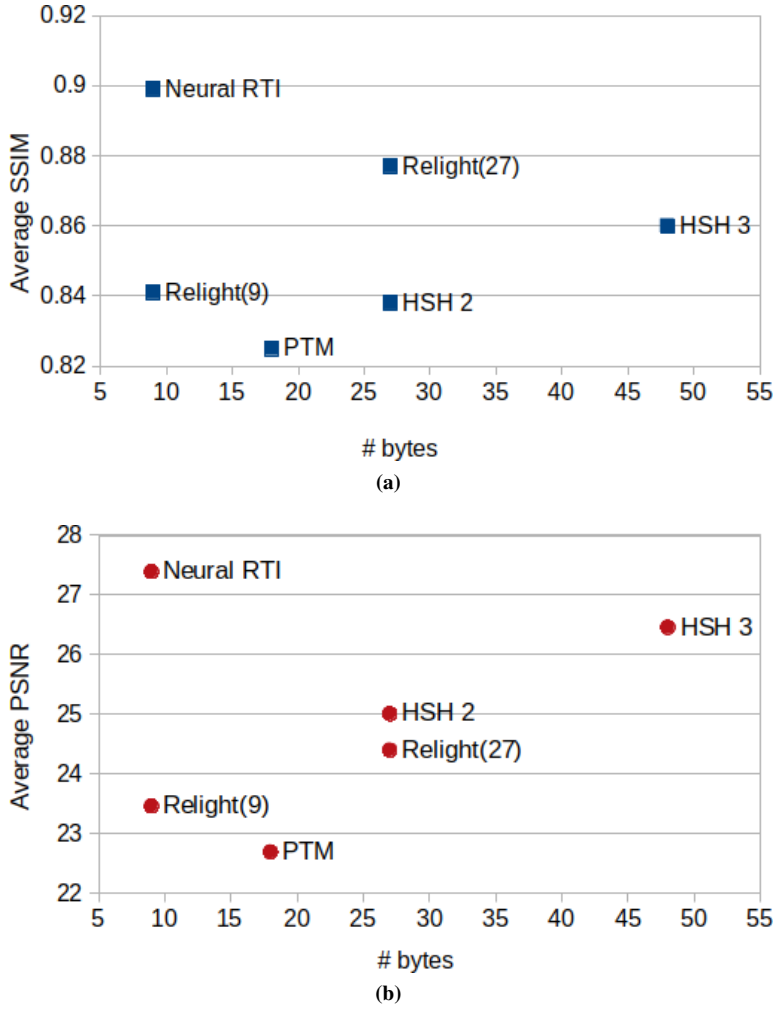


Figure 5.6: *Plots representing average SSIM (a) and PSNR (b) scores of the relighted images obtained with different methods from the test light directions of the MultiMaterial set. Scores are plotted versus the number of bytes used for the compressed encoding.*

network architectures. In [162], relighting is performed on 5 selected images, corresponding to those closer to an ideal sampling learned from a training set and to an encoding size of 5 bytes per pixel. This is, clearly, not the ideal way to compress a standard RTI dataset, but we still compared the relighting quality. Relighting is then global, based on the function learned on the training set. This can result in global illumination effects and realistic, even if hallucinated shadows.

The method in [120] is designed for a different application, e.g. compressing high dimensional multi-light and multi-view data of nearly planar patches. We instead focus on non-planar samples of potentially sharp BRDFs. The network of [120] uses 1D con-

	PTM	HSH(27)	HSH(48)	[116](9)	[116](27)	Neur.(9)
canvas	25.17	28.45	30.03	27.95	28.45	**33.26**
tablet	20.56	22.76	**24.24**	20.89	22.28	23.74
bas relief	22.34	23.81	25.10	21.54	22.47	**25.17**
average	22.69	25.01	26.46	23.46	24.40	**27.39**

Table 5.2: *Average PSNR of the methods on the different shapes (all MultiMaterial sets). Parentheses include the number of coding bytes.*

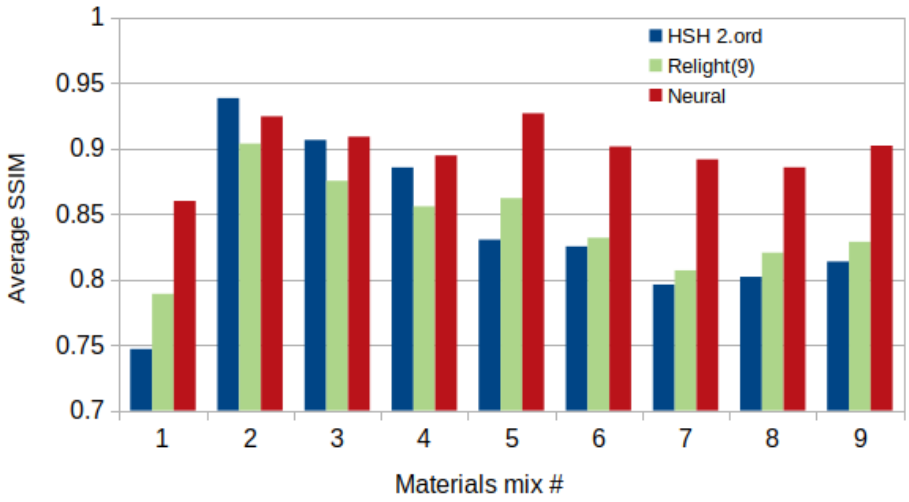


Figure 5.7: *Bar chart showing average SSIM of selected methods on the different materials. Neural relighting and 2nd order HSH perform quite similarly on rough materials, while neural relight is significantly better on glossy materials (5 and 6, see Table 3.1)*

volutions, while we avoid them and instead use more FC layers and different activation functions that work better with sharp-and-rotated BRDFs. We test an architecture similar to [120] in our case, with three 1D convolutional layers and a single fully connected layer, encoding with 9 bits and relighting similarly.

Looking at results (see Table 5.3), it is possible to see that the adapted methods do not provide good results when compared with our method and other RTI encodings, showing that a specifically designed neural architecture is the right choice for the RTI relighting task.

5.2.6 Ablation study

In our experiment, we have investigated the influence of different network parameters, i.e. the number of layers, number of units in each layer, different encodings as well as hyperparameters, specially on the decoder network. From our investigation, we found that increasing the number of layers does not change the performance that much whereas it increases the execution time. Above 3-4 hidden layers, the improvement in relighting

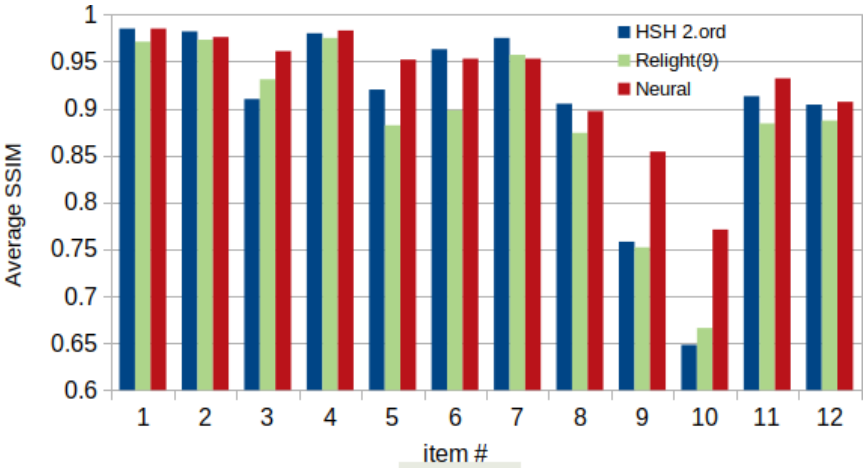


Figure 5.8: *Bar chart showing average SSIM of selected methods on the different objects. It is possible to see that the great improvement of the NeuralRTI relight is obtained for the shiny metallic objects (coins, items 9,10)*

accuracy compared to the increase in the size of the decoder is minimal. For this reason, we use a decoder network with 3 hidden layers as a compromise between accuracy and computation time. We have also investigated the effect of the number of units in each layer. As we can see from table 5.4 setting the number of units to 3N, where N(49 in this case) is the number of input lights gives us the best result. Despite setting the number of units to 3N is good from the same table we can see that reducing the units doesn't deteriorate the performance that much. However, it improves the execution time significantly.

Another study we have performed is the influence of different encodings. In this case, as shown in table 5.5, the performance slightly increases when we increase the encoding size. However, when we compare the performance gain to the execution time is almost insignificant. For this reason, we set the encoding size to 9.

5.2.7 Interactive relighting

We tested the Neural RTI encoding in an existing web-based solution for the interactive relighting RTI framework. The current encoding results in a relighting that is far more complex than the simple weighted sum of the RTI/HSH coefficients and involves several multiplications that are proportional to the squared decoder layer size, however, we kept the number and the size of the decoder layers sufficiently small to allow interactive relighting.

To implement the web viewer, we used the tensorflow.js library [134]. The 9 coefficients per pixel, quantized to 8 bits, are loaded in the browser as a binary typed array and mapped back to the original floating-point coefficients range. The light direction parameter, varying with the mouse position is concatenated with the coefficients and processed in the decoder network. Tensorflow.js library adopts a WebGL backend for the network processing and the data managed as textures in the GPU. The result is then rendered in a canvas element. Similarly to PTM, HSH and PCA web viewers approach the coeffi-

	Average PSNR		
	Ours (9)	Adapted Neural BTF (9)	Adapted Deep Relighting (15)
SynthRTI single	**31.00**	28.37	13.15
SynthRTI multi	**27.39**	23.49	12.46
Real	**32.31**	29.64	26.61
Average	**30.23**	27.16	17.41

	Average SSIM		
	Ours (9)	Adapted Neural BTF (9)	Adapted Deep Relighting (15)
SynthRTI single	**0.898**	0.852	0.449
SynthRTI multi	**0.899**	0.808	0.436
Real	**0.927**	0.89	0.836
Average	**0.908**	0.850	0.574

Table 5.3: *Adaptation of other network architectures not specifically designed for the relighting of typical RTI stacks resulted in poor results compared to our solution.*

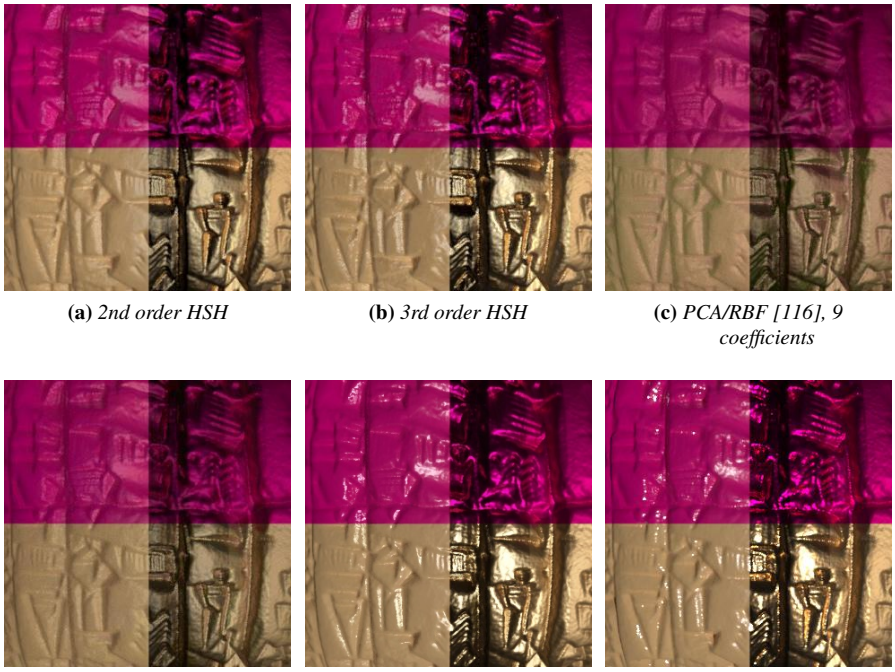
	Average SSIM			
	25 nodes	75 nodes	147 nodes	300 nodes
SynthRTI	0.841	0.842	**0.853**	0.844
Real	0.922	0.941	**0.947**	0.930
Average	0.881	0.891	**0.900**	0.887

	Average PSNR			
	25 nodes	75 nodes	147 nodes	300 nodes
SynthRTI	27.271	28.302	**28.744**	25.533
Real	32.148	32.210	**32.310**	31.102
Average	29.709	30.256	**30.527**	28.317

Table 5.4: *Image relighting quality of different network configurations. Here we can see that the network configuration with the number of nodes equal to 3N(in this case 147) at each layer performs well.*

	Average SSIM		
	Number of Nodes		
size of coefficient	100	147	200
9	0.8406	0.8414	0.8422
27	0.8404	**0.8437**	0.834
48	**0.8467**	0.8392	**0.8443**

Table 5.5: *Image relighting quality as a function of latent coefficient performed on bas-relief shape with the single materials. Here, we evaluate the quality of a relighted image by encoding in 9, 27 and 48 bytes.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure 5.9: *Comparison of images relighted from the encodings created with the Multi-Material 5 dome collection and the corresponding ground truth reference.*

cients can be combined and compressed in three JPEG or PNG images and split into a multiresolution pyramid of tiles.

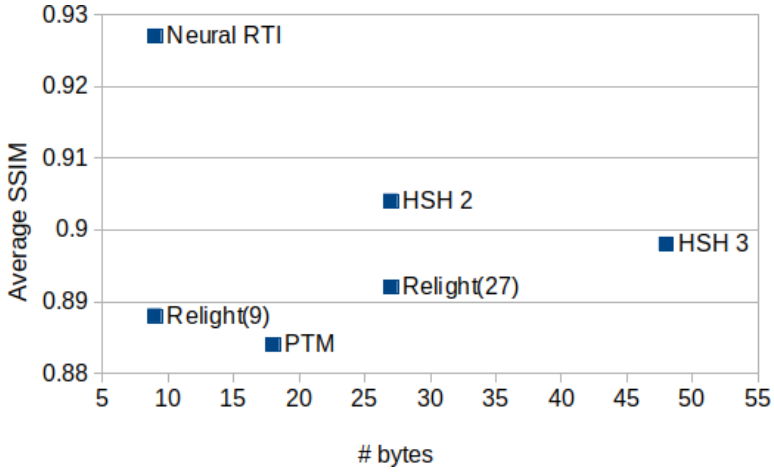
Figure 5.12 shows some snapshots from an interactive surface inspection performed with the web viewer on our encoding.

A gallery of interactively relightable web-based visualizations of a selection of SynthRTI and RealRTI items can be seen at the project webpage `https://univr-rti.github.io/NeuralRTI/`.

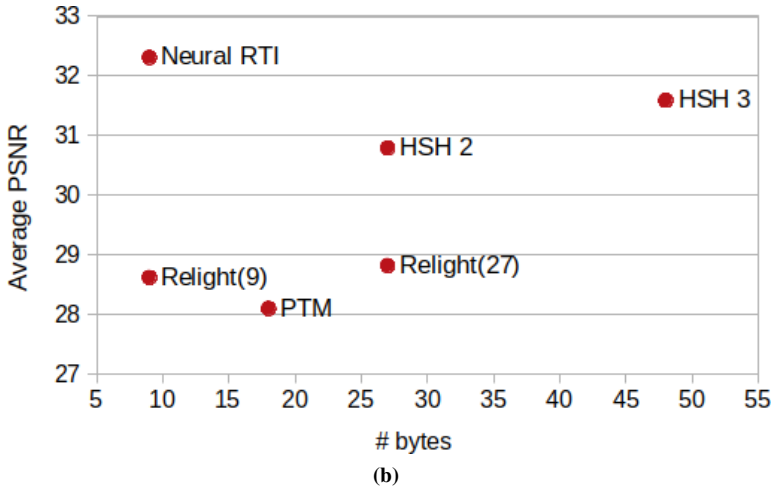
5.3 Summary

In this chapter, we presented a NeuralRTI, a neural network-based relighting tool that can provide better results than the current state-of-the-art methods, with reduced storage space. The quality of the images created with our novel technique is particularly good especially on surfaces with metallic and specular behavior that are not well handled by the existing methods. Furthermore, the relighted images are less affected by the blended shadows artifacts typical of RTI.

NeuralRTI can be used immediately as an alternative to the classic PTM, HSH and PCA-RBF files and could be directly integrated into the existing visualization frameworks,



(a)



(b)

Figure 5.10: *Average quality scores of relighted images on the RealRTI dataset (all 12 objects), as a function of number of bytes used for the encoding.*

as it is efficient enough to support real-time relighting. We demonstrate this with a specific web viewer based on tensorflow.js We believe that this tool will be particularly useful in the domain of Cultural Heritage and material science surface analysis, where RTI processing is widely employed.

The main limitation of the proposed technique is that it is a "local" method, not learning global effects like cast shadows. This limitation, however, holds similarly for the currently used techniques and it must be stressed that NeuralRTI provides better highlights and shadow simulations, avoiding blending artifacts. This is probably due to the ability to constrain non-linearly the space of the reflectance patterns.

We plan, as future work, to investigate how our method, developed to handle the typical

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *) NeuralRTI* **(f)** *ground truth image*

Figure 5.11: *Comparison of relighted images on a real challenging surface.*

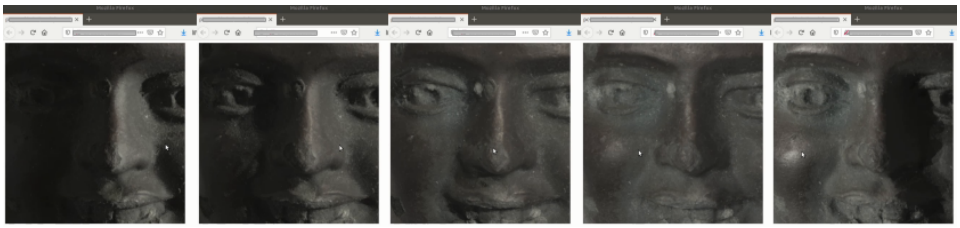


Figure 5.12: *Our web-based interactive relighting solution allows, like similar applications working with PTM/HSH/PCA-RBF encoding. Moving the cursor over the image, the user controls the light direction for the novel illumination. In this case, moving from right to left, he can gradually move from raking light from the right, to illumination from top, to raking light from left.*

sampling of RTI acquisition behaves with varying light directions sampling density, possibly specializing the codes for different acquisition protocols. Our compressed encoding can be also used as a basis for further surface analysis, e.g. material and shape characterization. We plan to test methods to recover effectively normals and BRDF parameters from the compact encoding of the captured objects.

CHAPTER 6

Crack Detection in Single- and Multi-Light Images of Painted Surfaces using Convolutional Neural Networks

Cracks represent an imminent danger for painted surfaces that needs to be alerted before degenerating into more severe aging effects, such as color loss. Automatic detection of cracks from painted surfaces' images would be therefore extremely useful for art conservators; however, classical image processing solutions are not effective to detect them, distinguish them from other lines or surface characteristics. A possible solution to improve the quality of crack detection exploits Multi-Light Image Collections (MLIC), which are often acquired in the Cultural Heritage domain thanks to the diffusion of the Reflectance Transformation Imaging (RTI) technique, allowing a low cost and rich digitization of artworks' surfaces. In this chapter, we propose a pipeline for the detection of crack on egg-tempera paintings from multi-light image acquisitions and that can be used as well on single images. The method is based on single or multi-light edge detection and on a custom Convolutional Neural Network able to classify image patches around edge points as crack or non-crack, trained on RTI data. The pipeline can classify regions with cracks with good accuracy when applied on MLIC. Used on single images, it can give still reasonable results. The analysis of the performances for different lighting directions also reveals optimal lighting directions.

6.1 Introduction

Cracks are damage that can affect Cultural Heritage objects on many layers. For instance, as pointed out in [29], cracks can affect not only the pictorial layer of a painting but the varnish and support layers as well. Cracks may occur due to several reasons: the drying of the paint layer (since the evaporation of organic components causes their shrinkage), external mechanical factors (vibrations, seism, impacts) and stress induced by aging or fluctuations of humidity over time [115], where the cracks burst as non-uniform contractions from the substrate layer through the superficial layer [31]. Hence, cracks represent a form of degradation both aging-dependent and aging-independent and the detection of such degradation at any stage in the lifecycle of artwork reveals meaningful clues for conservators. Moreover, in the case of paintings, the craquelure (which is the network of connected cracks) might also be associated with a certain school of painting and can be useful for recognizing the style of an artist [12, 13]. Cracks are multi-surface phenomena, since, apart from paintings, they have been investigated in concrete surfaces as well [21, 86]. In addition to this, cracks are often used for structural health monitoring of cultural heritage buildings, where commonly mechanical displacements lead to deep discontinuities in structural elements of the building such as ancient support walls [75].

In this chapter we propose a novel approach for crack detection exploiting both MLIC typically acquired in the CH domain and Convolutional Neural Networks. The idea is to exploit available multi-light image data of aged egg-tempera painting to develop a pipeline for crack detection in similarly painted surfaces based on (multi-image) edge detection and a custom CNN based classifier to select crack candidates and label them in a supervised manner.

The contribution of the work is both the demonstration of a practical pipeline that can be directly used on similar surfaces or retrained for other surface feature detection, analysis of the improvements of the edge detection and classification approaches with MLIC data.

The chapter is structured as follows. In Section 2, we review the related works. In section 3, we present the rationale of the method, the dataset used and describes all the processing steps. Finally, Section 4 presents the experimental results and section 5 concludes the chapter.

6.2 Related Work

Automatic detection of cracks on digital images, especially in the case of painted surfaces, is particularly difficult, as the geometric and color features visually characterizing them are quite subtle and difficult to discriminate from those related to drawing lines, noise, and other aging effects.

The common pipeline for cracks classification in digital images is therefore complex, and usually involves a step of pre-processing consisting of morphological operators: white and black top-hat transformations [29], opening, closing, spurring and cleaning [21]. After the pre-processing step, the classification of cracks proceeds either with an unsupervised, supervised, or semi-supervised approach. In the unsupervised approach, usually, a combination of edge detection and further heuristics are used to label cracks [31].

To restore the Ghent Altarpiece, [127] proposes a crack detection workflow followed by crack inpainting. For reducing the noise in the image, they initially apply anisotropic

diffusion filtering and proceed with crack detection based on a multiscale morphological approach by switching the structuring element within the top-hat transform to different sizes. Besides, they explore the distinct contrast provided by the color channels of RGB and HSV color spaces either for identifying cracks with extreme brightness values or for isolating the misidentified cracks. Thus, they found that the green channel is enhancing the dark cracks, while the blue channel enhances the bright cracks. Similarly, they were able to distinguish between deceivingly crack-resembling elements (brows) and dark cracks, by applying heuristic on the saturation channel of the HSV image. As an extension of the work in [127], in [26] two other methods are proposed except the multiscale top-hat transform: the oriented elongated filters originally applied for blood vessels segmentation and the k-SVD dictionary learning with hysteresis thresholding. To validate these methods, a semi-automatic clustering is performed with a k-means algorithm that receives as input a feature vector composed of several joint color and shape properties of the crack pixels and their neighborhood. The latter is particularly descriptive of the bright borders that usually surrounds cracks. In [115], built on top of [127] and [26], a semi-supervised approach is adopted for crack detection in multimodal images (visible, infrared and x-ray radiography). The rationale of their improved method lies in using a Bayesian conditional tensor factorization (BCTF), by estimating for each multimodal pixel the posterior probability of pertaining to the "cracks" class.

Nonetheless, in the case of crack detection in paintings, [31] suggest that a supervised approach is highly recommended because there are elements within a painting (especially thin and dark brushstrokes on a bright background or bright brushstrokes on a dark background) that visually resemble the structure of a crack, so the unsupervised algorithm outputs increased false positives.

Among the supervised learning approaches applied in the image processing domain, the most popular now is, clearly, the use of Convolutional Neural Networks (CNN). CNN's have been used to solve the problem of automatic crack detection in concrete surfaces [16]. In [119], CNNs are combined with 3D modeling for spotting defects in tunnel infrastructures. In [94], authors improve a standard genetic algorithm by training a CNN, with only a minor increase in cracks' classification accuracy for general-purpose tasks.

Another interesting option for better crack detection is the use of Multi-Light imaging, aka Photometric Stereo or Reflectance Transformation Imaging. This kind of image acquisition, consisting in taking multiple photos from a single viewpoint with changing light direction is quite popular in the Cultural Heritage domain as it can be obtained with low-cost setups and provides an effective visualization of surfaces [106]. Multi-Light image collections (MLIC) have been used for crack detection [66, 70, 82, 133, 136], typically using few lights and specific setups.

6.3 The proposed approach

To develop and validate crack detection algorithms based on Multi-Light Image Collections (MLIC), we rely on a set of acquisitions obtained from a European project on the analysis of artworks' aging, Scan4Reco [34]. In this project, mockups of wood pantings have been realized using different pigments and coatings and then artificially aged to characterize degradation effects. All the items have been acquired at different aging steps with a free-form RTI setup, with the acquisition and processing pipelines presented in [23].

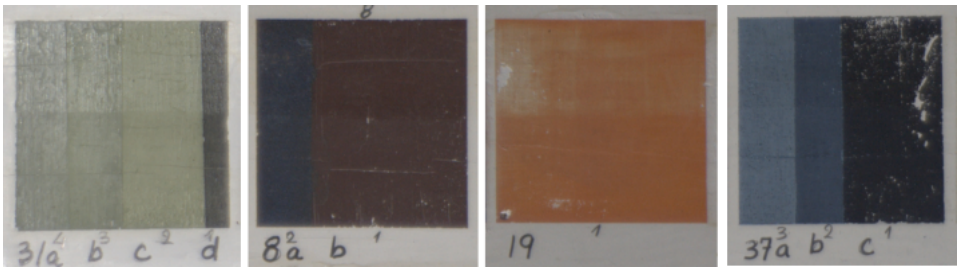


Figure 6.1: *Artificially aged egg-tempera samples. Images are taken from the original MLIC captures featuring 50 different light directions.*

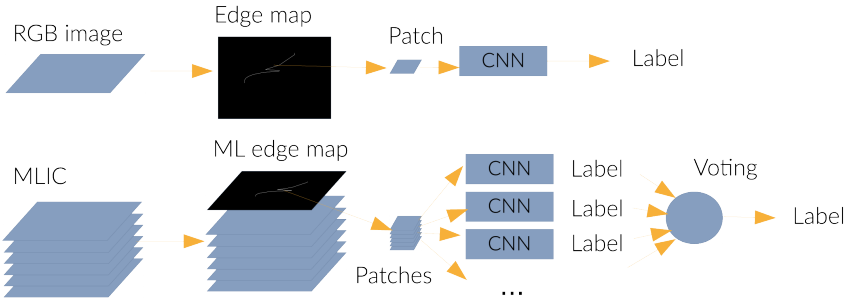


Figure 6.2: *The proposed crack detection pipeline: edges are extracted on single images or MLICs and single or multiple patches around each edge location are classified with the same CNN-based classifier. In the case of MLIC based crack detection, edge labels corresponding to the different patches are combined by majority voting.*

The calibration procedures applied results in a set of 50 intensity-compensated images, with estimated light direction [48].

Our idea is to exploit the multi-light images, captured at the second aging step representing 36 painted squares with different pigments and coatings, and presenting visible cracks that can have been annotated by experts to develop and test a novel approach to automatically identify regions with cracks on egg-tempera paintings.

The proposed method is based on a processing pipeline able to automatically identify crack regions in single and multi-light images. The pipeline is based on a (multi-light) edge detection and a Convolutional Neural Network-based labeling of image patches around edges. The processing pipeline is summarized in Figure 6.2.

In the following subsections, we present the rationale of the different steps.

6.3.1 Edge detection and cracks

Figure 6.3 shows that the detection of the cracks identified in these images by CH experts is quite hard. The top row show two images of a MLIC of cracked egg tempera and the second edge maps extracting thresholding Sobel gradient magnitude estimation. Part of the cracks are not visible in both images and consequently in the edge maps. The complete

detection of the cracks is possible looking at the whole dataset and it is reasonably captured by multi-light edge detection.

A few methods have been proposed to extract edges from MLIC/RTI data. For example Brognara et al. [11] localize edges as maximal variations of 3D normals computed from Polynomial Texture Map fitting. Pan et al [100] also use PTM coefficients, but estimate edges using the idea of Di Zenzo [33] for multidimensional image edge detection, e.g. estimating the Jacobian matrix and evaluating eigenvalues to determine gradient magnitude and direction.

In our work we just applied the Jacobian approach on the set of the intensity images of the MLIC collection to recover a gradient intensity, that is thresholded in a conservative way to recover a superset of the edge points that are considered as candidate crack locations. An example result of the procedure is represented in the left image of the bottom row of Figure 6.3.

6.3.1.1 Ground truth crack annotation

The annotation of the ground truth crack position has been performed as follows: experts were provided with selected images of the MLIC where cracks were maximally visible and the edge map provided by the multi-light edge detector. They had to draw a polygonal area including the edges that should be classified as a crack. The annotation resulted in binary maps corresponding to each of the 36 MLIC used.

6.3.2 CNN-based edge classification

To automatically label detected edges as crack or non-crack, we consider a $31x31$ patch around them and train a classifier based on a Convolutional Neural Network (CNN). CNNs [50] are powerful learning tools demonstrating superior performance on both visual object recognition and image classification tasks [68]. As the number and the direction of the input lights are not necessarily the same in different acquisitions, we designed a classifier predicting the feature class (crack/non-crack) based on a single image patch. After the training process, the classifier can then be used both to classify features extracted on a single image or to classify features extracted on a multi-light acquisition using a voting approach: given an edge point extracted on a multi-light image collection, we apply the classifier to all the patches of the images centered in the point and assign the most frequent label.

6.3.2.1 Patch dataset creation and labelling

To train the classifier, we considered a subset of the ground truth crack pixels (avoiding to take close pixels that would result in heavily correlated patches) and, for each MLIC, an equal number of positive(crack) and negative(noncrack) samples, Figure 6.4.

To evaluate the patch classification method and the full pipeline, we divided the dataset, using 30 MLICs of painted squares and the corresponding annotated patches as a training set, and the remaining 6 painted squares and the corresponding patches as a test set, both for the classifier evaluation and the full crack detection pipeline testing.

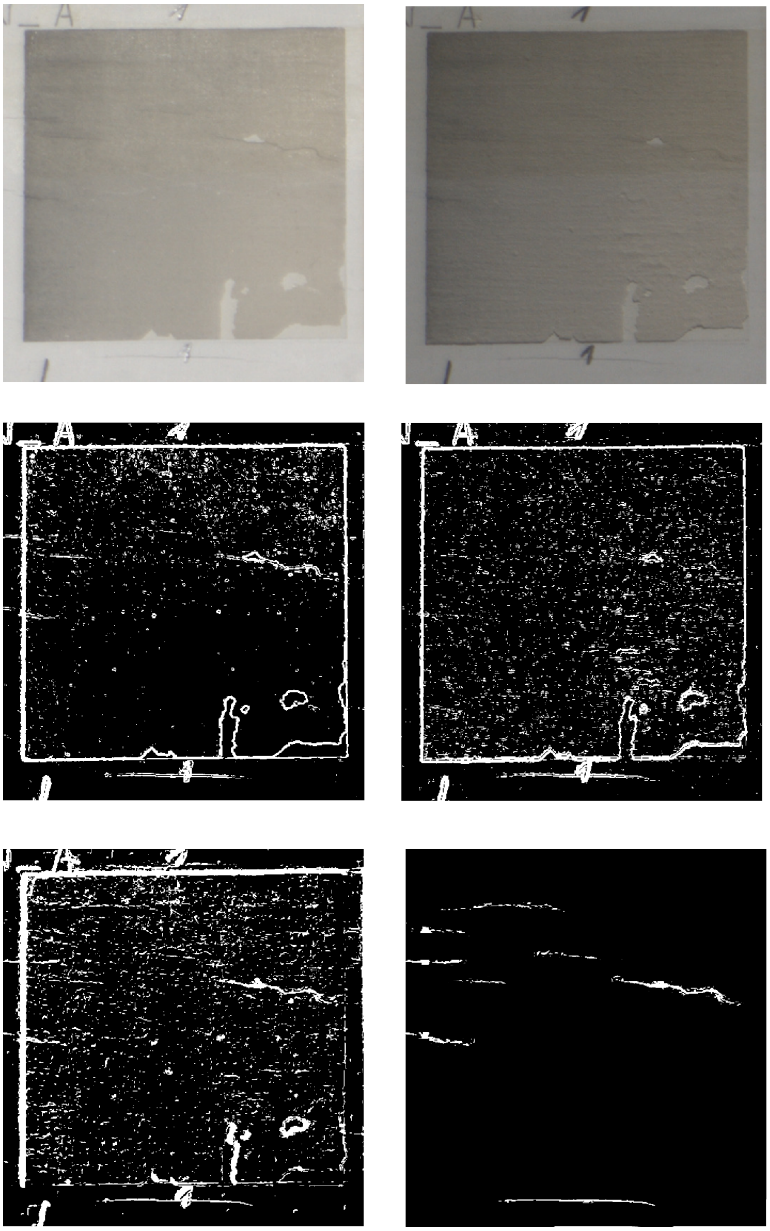


Figure 6.3: *Top row: two images of an egg tempera sample MLIC. The visibility of the surface features changes with respect to the light direction i.e. more visible on the high elevation and less visible on the low elevation. Middle row: the corresponding edge maps estimated on the single images. Bottom row: MLIC based edge detection (first column) and edge points annotated as cracks (second column).*

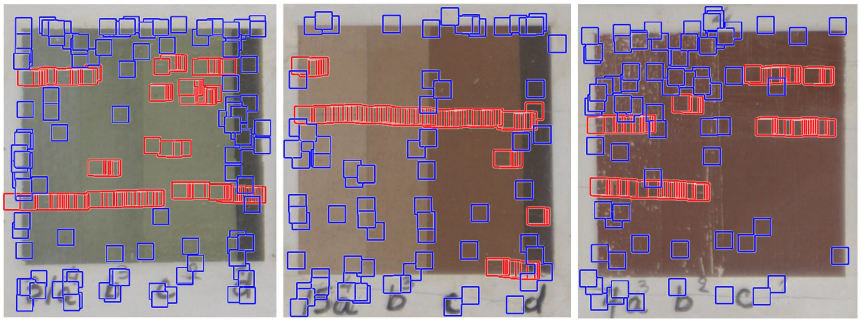


Figure 6.4: *Examples of positive (red) and negative (blue) patches centered in edge points and used to train (and test) the classification of candidate crack points.*

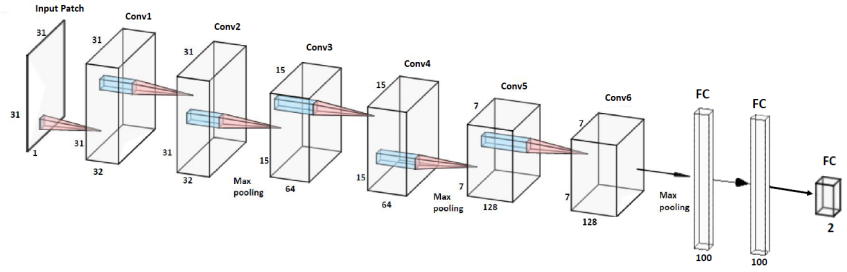


Figure 6.5: *An illustration of the architecture of our CNN. Conv#: layers corresponding to convolution operations and max-pooling: max pooling applied on the previous layer.*

6.3.2.2 Network architecture

Figure 6.5 shows our CNN model. It contains six convolutional layers and three fully connected layers. The first two fully-connected layers have 100 neurons each and the last one contains 2 neurons. The input is a 31x31 grayscale patch, like those obtained with the procedure described in Section 6.3.2.1). In the first layer, the input patch is convolved with 32 filters of size 3x3, outputting 32 feature maps of size 31x31 each. In the second layer, the same operation is performed followed by 2x2 max-pooling and subsampling by a factor 2. In the third and fourth layers, the feature maps are convolved with 64 3x3 filters and followed by 2x2 max-pooling and subsampling, resulting in 64 7x7 feature maps. In the fifth and sixth layers, input features are convolved with 128 3x3 filters and followed by max-pooling. The output of the last fully connected layer is fed to a 2-way softmax which produces a distribution over the 2 class labels. All hidden layers are equipped with the Rectified Linear Unit (ReLU) activation function.

The model is trained on 54,810 training samples which contain an equal number of positive and negative samples and validated on 6,087 samples. It is trained using Adam optimization algorithm [65] with a batch size of 64 examples, a learning rate of 0.0003, a Gradient decay factor of 0.9 and a squared gradient decay factor of 0.99. We found that this combination of parameters was important for the model to learn. Batch normalization and dropout layers [137], which can prevent overfitting, with a rate of 0.2 (20% dropout)

are also used. It is trained on a GeForce RTX 2080 Ti machine with a single GPU.

In our tests, this architecture seemed better suited for the task with respect to several other CNNs tested, like other deep networks originally developed for crack detection in concrete [73] or LeNet5 [71] which has an architecture with a lower number of convolutional layers and a lower number of filters. The network we adopted is a reasonable tradeoff between complexity and trainability.

6.3.3 Full detection pipeline

Using the trained classifier, the idea of the method is to use the proposed pipeline to automatically extract crack points as follows: if the input is a single image, a Sobel edge detection is performed using conservative thresholding and then the candidate points are classified. Finally, the map of positive points is post-processed by removing isolated points. In the case of MLIC data, the edge map is extracted with the Di Zenzo like multi-light edge detection, for all the pixels we classify all the corresponding patch and obtain the final label with majority voting. Finally, we post-process the map removing isolated points.

6.4 Results

In our experiments, we both evaluated the performances of the patch classification on the dataset created from the Scan4Reco painted samples and evaluated the use of the detection pipeline to automatically detect cracks.

6.4.1 Patch classification

To evaluate the quality of the edge classification, we evaluated the classification accuracy on the annotated dataset patches extracted from the six test squares. Table 6.1 shows the classification errors for the patches of the different MLIC test data obtained with single image classification and MLIC classification based on the same CNN-based classifier and majority voting. The accuracy is not very high, but the task is quite hard due to the differences in pigments and coatings used in the different squares. The use of MLIC data makes the classification more accurate as expected. If we look at the results obtained with single images, however, we see that the accuracy is not that bad, and depends on the light direction. If we consider the variation in classification accuracy versus the elevation of the light used, it is possible to see (Table 6.2) that using perpendicular lights the automatic classification works better (this may be counter-intuitive, as a typical way of inspecting surfaces changing light direction is the use of raking lights.

6.4.2 Automatic crack detection

To automatically detect cracks from single light images it is, however, necessary also to rely on single image edge detection, that is less able to recover all the correct candidate crack points to be classified.

Figure 6.6 shows the cracks detected on six test squares. Results seem good despite the quite hard visibility of the cracks and the huge amount of clutter, removed during post-processing using a simple morphological operator, area opening.

Test square	Single Image	Voting
1	65.64	72.26
2	81.33	90.08
3	85.24	90.49
4	87.11	87.85
5	75.01	75.14
6	66.65	76.87
Avg.	**76.83**	**82.12**

Table 6.1: *Classification accuracy for the dataset patches extracted on the 6 test painted squares. We report average accuracy obtained on single patch classification and the average accuracy on voting based MLIC-based classification*

	Accuracy (%) in elevation range		
	0-30 deg.	30-60 deg.	60-90 deg.
1	60.75	64.63	68.92
2	77.01	82.39	84.60
3	81.31	87.56	86.84
4	85.63	87.88	87.83
5	67.87	77.96	81.77
6	58.96	66.11	74.88
Avg.	**71.92**	**77.75**	**80.81**

Table 6.2: *Average accuracy in single patch edge classification for selected elevation ranges. Cracks are better recognized when the illumination is from higher elevation angles.*

The results obtained on the multi-light data (middle row) appear better than those obtained on the single image example, that is, however, rather good. This is due to the better quality of both the edge detection performed with the multi-light Di Zenzo approach and the better classification accuracy obtained using CNN.

As can be seen in figure 6.7, in the case of prediction on single images, the quality appears better on high elevations despite the middle and low elevations are also providing satisfactory results. If we see the crack map lines superimposed over the image(cyan points), on the high elevations(right image), the cracks' maps lines are continuous. This continuity shows in most of the cases that the classifier can classify the crack points correctly. Probably the increased illumination intensity and the subsequent inter-reflections are more effective than raking light for the enhancement of the crack details.

To demonstrate the crack detection on real paintings, we used our method on a MLIC scan of a MLIC capture of real artwork, e.g. the Icon St. Michele (17^{th} - 18^{th} century). It is an egg tempera painting on wood support and includes some regions with cracks. As can be shown in Figure 6.8, it is possible to see that the proposed method can extract reasonably well the cracks neglecting most of the other image edges not corresponding to cracked painting.

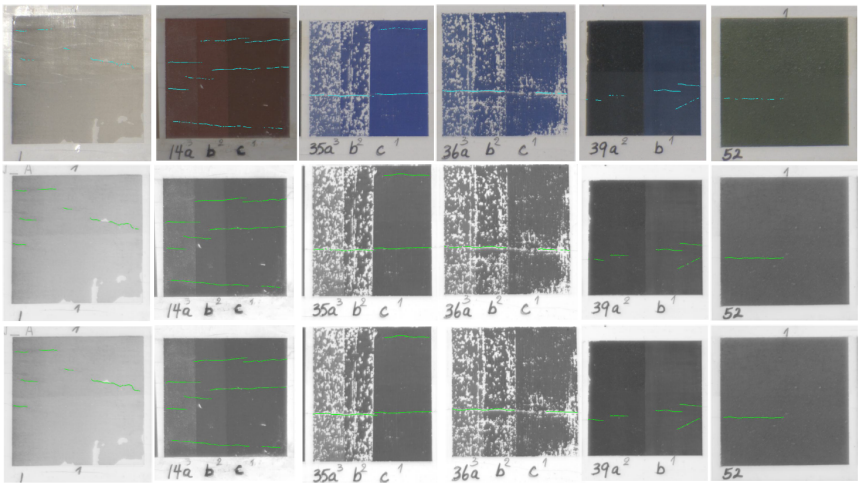


Figure 6.6: *Cracks detection on test painted squares. Top row: corresponding cracks detected on a single image after CNN classification superimposed on images (cyan points). Middle row: corresponding cracks detected on MLIC after CNN classification, superimposed on albedo image (green points). Third row: ground truth crack points superimposed on albedo image (green points).*

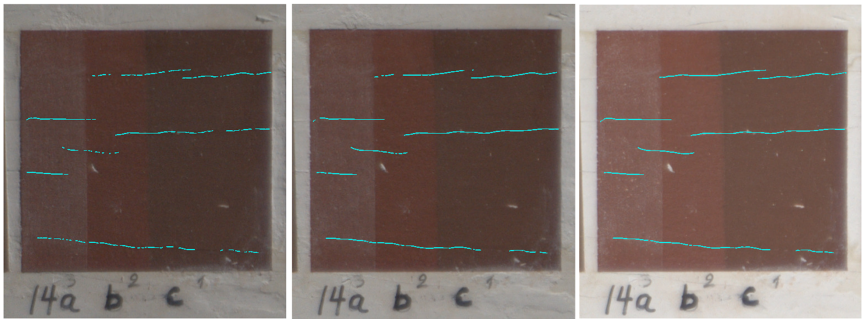


Figure 6.7: *Crack points detected on a single image at various elevation angles superimposed on the corresponding images. Left to right crack detected from an image captured at elevation angles of 18, 44 and 66 respectively. As we can see, on the left image the lines which represent a crack edge are disconnected. Whereas in the middle and mainly on the right one they are connected. This tells us that on the high elevation almost all the crack points are detected precisely and on the low elevation not.*

6.5 Conclusion

The automatic identification of specific painting features like cracks can be extremely useful for conservators to monitor the aging of the items, highlighting possibly critical regions of the surface.

The use of machine learning tools coupled with multi-light imaging is surely a good

Figure 6.8: *Crack points detected on real images. Left: example image chosen from the Multi-light image collection, one of the few where cracks are visible. Middle: grayscale albedo. Right: corresponding cracks detected on MLIC after CNN classification, superimposed on the grayscale albedo image (green points).*

way to address the problem of automatic crack detection, but it is not widely studied in the literature. We proposed a specifically designed pipeline that can be used to detect cracks (but also other critical features with specific training). Results are encouraging, even if they could be certainly improved and tested on more data with heterogeneous features.

One problem related to the use of machine learning for the automatic interpretation of CH data is the lack of large annotated databases specifically designed to solve practical problems for analysis and conservation applications. Therefore, We plan to collect novel datasets to evaluate the proposed approach on larger collections of MLIC data.

CHAPTER 7

Conclusion

Multi-Light Image Collections(MLICs) are a series of images, acquired from the static viewpoint but varying the lighting condition, arranged in a stack. They are used as the main means to non-destructively gather information on scenes and objects at many scales, as well as to provide users with useful visualization tools for object analysis. In this thesis, we present methods that exploit Multi-Light Image Collections (MLICs) for surface analysis and visualization and proposed schemes that exploit MLIC to create a relightable image and to attack the problem of crack detection. Moreover, we have proposed benchmark datasets and evaluated the performance of different photometric stereo algorithms for cultural heritage applications.

The first chapter provides background information on Multi-light image collections. Here, we discuss different acquisition setup used by the MLIC community, categorizing them into two main categories. Then, we discuss different light calibration techniques and conclude the chapter discussing application areas where MLICs have been successfully used for the research of daily analysis work.

In chapter 2, we discuss the use of MLIC for surface visualization and analysis and point out available tools used to support the analysis. In this chapter, we discuss methods that strive to support the direct exploration of the captured MLIC, that generate relightable models from MLIC, non-photorealistic visualization methods that rely on MLIC and visualization tools used to do MLIC analysis.

One of the challenges in the Multi-Light Image Collections(MLIC) community, is the lack of benchmarks for evaluation. Usually, MLIC based methods evaluated using specific homemade datasets or a few publicly available benchmark(for example DiLiGent,

[132]). The main problem of these types of a benchmark is that they are not representative of those typically captured in the real-world applications of MLIC acquisition. Thus, the availability of specific benchmarks for different surface analysis tasks will be fundamental to validate methods and develop guidelines for the use of MLIC visualization software in different domains. To this end, in chapter 3 we have proposed benchmarks(RealRTI, SynthRTI and SynthPS) that can be used to evaluate algorithms that rely on MLIC. Utilizing RealRTI and SynthRTI we have evaluated the NeuralRTI method. In chapter 4, we have evaluated the performance of different photometric stereo algorithms using SynthPS.

In chapter 5, we proposed a neural network-based RTI method, aka NeuralRTI, a framework for pixel-based encoding and relighting of RTI data. In this method using a simple autoencoder architecture, we show that it is possible to obtain a highly compressed representation that better preserves the original information and provides increased quality of virtual images relighted from novel directions, particularly in the case of challenging glossy materials.

Finally, in chapter 6 we proposed a method for the detection of crack on the surface of paintings from multi-light image acquisitions and that can be used as well on single images. The proposed method is based on a processing pipeline able to automatically identify crack regions in single and multi-light images. The pipeline is based on a (multi-light) edge detection and a Convolutional Neural Network-based labeling of image patches around edges.

Challenges for future research

Finally, we want to highlight promising and challenging research areas for future work. We summarize the main identified ones in the following sections.

Standardization of protocols, data exchange formats, and visualization approaches

One relevant challenge for the MLIC data visualization side is the development of standardized acquisition, processing, encoding, and visualization methods for this kind of data. Most image capture setups share a light-direction based parameterization, even if with different densities and constraints. However, there are no well-defined standards to encode the calibrated data stack. Moreover, in all the intermediate steps of the MLIC processing pipeline (e.g. light calibration, data storage, reflectance models, and parameters encoding) there is a clear lack of guidelines and widely accepted exchange formats. Also, for the relightable image visualization, even if the requirements for most practical applications seem to be quite similar (users need to be able to relight the surfaces with novel illumination, using false colors to represent attributes over the surface, visualize material signatures clicking on specific points), there are not standard container formats for the data that allows for interoperability of different solutions. The RTI framework [20] can be considered a de facto standard for some applications in CH since it is based on free and known software, but many more recent and performing solutions for relightable images and enhancements are not integrated and not compatible. Moreover, creators of light dome applications typically develop proprietary solutions for encoding and rendering data. It would be therefore useful soon to spend some efforts on the development of standardized protocols and data formats. This would also make it possible to experiment with novel visualization and user interface techniques.

Improving direct visualization

The improvement in capture techniques is making dense acquisition more practical. Direct visualization techniques (Sec. 2.1) have shown their applicability in several domains, but are currently under-developed. The need to access at visualization time large amounts of data has, in particular, restricted their applicability, especially in remote settings, where bandwidth is limited. Relighting based on low-frequency fitting (Sec. 2.3) is by far the most commonly employed visualization, but significant inaccuracies have been shown to exist [105]. Early attempts at using resampling and compression methods to support direct interpolation are promising [116], but this research is in the early stage and only supports aggressive lossy compression. It is reasonable to expect that the area of direct visualization can be significantly improved, tackling at least two different directions. On one hand, the interpolation of nearby images can be improved by exploiting more prior knowledge instead of using plain smoothness constraints. Moreover, multiresolution structures and compression techniques can be exploited, as done, e.g., in massive volumetric rendering [7], to allow for the real-time browsing of full MLICs on bandwidth- and storage-limited settings.

Smart use of machine learning techniques for relighting

In chapter 5 we have shown that using an artificial neural network, aka deep learning, it is possible to perform RTI, image relighting, better than the traditional model-based method. Deep learning is a powerful learning method inspired by how the brain works. For MLIC processing aimed at recovering shape (Photometric Stereo), as demonstrated in chapter 4 CNN-based methods have recently demonstrated the possibility of recovering normals better than traditional methods. So, it is expected that these methods could be used also to recover other intrinsic properties of the material for visualization. A key factor for the development of novel machine learning solutions will consist in the creation of large datasets of real and simulated MLIC datasets to train the algorithms. Accurate and fast rendering techniques will be required to quickly generate images, but also, for example, to compute loss functions in CNN training based on a similarity between rendered and reference images.

Improved interaction and data fusion for illustrative and perceptually-motivated presentation

While a variety of data fusion and illustrative techniques have been presented, this area seems under-developed to other visualization areas, such as, e.g., volumetric data exploration, where these techniques have shown to significantly improve data understanding. Several data-fusion techniques have been introduced, e.g., in multispectral/hyperspectral-only data-fusion and visualization [163] that could be explored also to enhance the perception of details in MLICs. Moreover, besides few examples (Sec. 2.4), most of the used techniques perform just overlays of feature maps and/or contrast enhancement using manipulations of some shape and appearance parameters. User interaction, in these cases, remains very limited.

In the area of visualization, however, many promising solutions have been explored, including a vast array of perceptually-motivated illustrative rendering approaches [118], and several interactive tools connected to illustrative methods, such as interactive lenses [143].

We foresee as a promising research direction the integration and further development of these techniques into the MLIC visualization and analysis pipeline.

Availability of specific benchmarks and datasets

Despite we have introduced the MLICs benchmark dataset (chapter 4), we believed that a more benchmark dataset, that can consider different material behaviour, is highly needed for future evaluation of different MLIC based algorithms. As previously shown, the usefulness of the different visualization approaches is not easily evaluated with a simple quantitative comparison of relighted images or normal maps with a ground truth reference by using common similarity metrics. From a visualization point of view, we should be able to evaluate the possibility of performing different tasks, e.g., searching for specific information, or evaluating details without distortion or artifacts. The task should be performed on different materials in accordance with the end user applications. Thus, the availability of specific benchmarks for different surface analysis tasks will be fundamental to validate methods and develop guidelines for the use of MLIC visualization software in different domains.

Bibliography

[1] Jens Ackermann, Simon Fuhrmann, and Michael Goesele. Geometric point light source calibration. In *Proc. VMV*, pages 161–168, 2013.

[2] Jens Ackermann and Michael Goesele. A survey of photometric stereo techniques. *Foundations and Trends in Computer Graphics and Vision*, 9(3-4):149–254, 2015.

[3] Jahanzeb Ahmad, Jiuai Sun, Lyndon Smith, and Melvyn Smith. An improved photometric stereo through distance estimation and light vector optimization from diffused maxima region. *Pattern Recognition Letters*, 50:15–22, 2014.

[4] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[5] Maria E Angelopoulou and Maria Petrou. Uncalibrated flatfielding and illumination vector estimationfor photometric stereo face reconstruction. *Machine vision and applications*, 25(5):1317–1332, 2014.

[6] Paula Artal-Isbrand and Philip Klausmeyer. Evaluation of the relief line and the contour line on greek red-figure vases using reflectance transformation imaging and three-dimensional laser scanning confocal microscopy. *Studies in Conservation*, 58(4):338–359, 2013.

[7] Marcos Balsa Rodriguez, Enrico Gobbetti, José Antonio Iglesias Guitián, Maxim Makhinya, Fabio Marton, Renato Pajarola, and Susanne Suter. State-of-the-art in compressed gpu-based direct volume rendering. *Computer Graphics Forum*, 33(6):77–100, September 2014.

[8] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.

[9] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[10] Simon Brenner, Sebastian Zambanini, and Robert Sablatnig. An investigation of optimal light source setups for photometric stereo reconstruction of historical coins. In *Proc. GCH*, 2018.

[11] Cristian Brognara, Massimiliano Corsini, Matteo Dellepiane, and Andrea Giachetti. Edge detection on polynomial texture maps. In *International Conference on Image Analysis and Processing*, pages 482–491. Springer, 2013.

[12] Spike Bucklow. The description and classification of craquelure. *Studies in conservation*, 44(4):233–244, 1999.

[13] Spike L Bucklow. A stylometric analysis of craquelure. *Computers and the Humanities*, 31(6):503–521, 1997.

[14] Martin D Buhmann. *Radial basis functions: theory and implementations*, volume 12. Cambridge university press, 2003.

[15] Moshe Caine and Michael Magen. Pixels and parchment: The application of RTI and infrared imaging to the Dead Sea scrolls. In *EVA*, 2011.

[16] Young-Jin Cha, Wooram Choi, and Oral Büyüköztürk. Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.*, 32(5):361–378, May 2017.

[17] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. Self-calibrating deep photometric stereo networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8739–8747, 2019.

[18] Guanying Chen, Kai Han, and Kwan-Yee K. Wong. Ps-fcn: A flexible learning framework for photometric stereo. In *ECCV*, 2018.

[19] Tongbo Chen, Michael Goesele, and H-P Seidel. Mesostructure from specularity. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1825–1832. IEEE, 2006.

[20] CHI. Cultural heritage imaging website, 2019. [Online; accessed-March-2019].

[21] G. K. Choudhary and S. Dey. Crack detection in concrete surfaces using image processing, fuzzy logic, and neural networks. In *2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)*, pages 404–411, October 2012.

[22] Hin-Shun Chung and Jiaya Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[23] I. Ciortan, R. Pintus, G. Marchioro, C. Daffara, A. Giachetti, and E. Gobbetti. A practical reflectance transformation imaging pipeline for surface characterization in cultural heritage. In *Proc. GCH*, pages 127–136, Goslar Germany, Germany, 2016. Eurographics Association.

[24] I.M. Ciortan, T.G. Dulecha, A. Giachetti, R. Pintus, A. Jaspe-Villanueva, and E. Gobbetti. Artworks in the spotlight: characterization with a multispectral led dome. In *IOP Conference Series: Materials Science and Engineering*, volume 364, page 012025. IOP Publishing, 2018.

[25] Toby Collins and Adrien Bartoli. 3D reconstruction in laparoscopy with close-range photometric stereo. In *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 634–642. Springer, 2012.

[26] B. Cornelis, T. Ruzić, E. Gezels, A. Dooms, A. Pizurica, L. Platisa, J. Cornelis, M. Martens, M. De Mey, and I. Daubechies. Crack detection and inpainting for virtual restoration of paintings: The case of the Ghent Altarpiece. *Signal Processing*, 93(3):605–619, March 2013.

[27] Massimiliano Corsini, Marco Callieri, and Paolo Cignoni. Stereo light probe. In *Computer Graphics Forum*, volume 27, pages 291–300. Wiley Online Library, 2008.

[28] Antonino Cosentino, Samantha Stout, and Carmelo Scandurra. Innovative imaging techniques for examination and documentation of mural paintings and historical graffiti in the catacombs of San Giovanni, Syracuse. *International Journal of Conservation Science*, 6(1), 2015.

[29] H. Deborah, N. Richard, and J. Y. Hardeberg. Hyperspectral crack detection in paintings. In *2015 Colour and Visual Computing Symposium (CVCS)*, pages 1–6, August 2015.

[30] Matteo Dellepiane, Massimiliano Corsini, Marco Callieri, and Roberto Scopigno. High quality PTM acquisition: Reflection transformation imaging for large objects. In *Proc. VAST*, pages 179–186, 2006.

[31] S. D. Desai, K. V. Horadi, P. Navaneet, B. Niriksha, and V. Siddeshvar. Detection and Removal of Cracks from Digitized Paintings and Images by User Intervention. In *2013 2nd International Conference on Advanced Computing, Networking and Security*, pages 51–55, December 2013.

[32] DHLAB. RTI tools at DHLAB Basel, 2019. [Online; accessed-March-2019].

[33] Silvano Di Zenzo. A note on the gradient of a multi-image. *Computer vision, graphics, and image processing*, 33(1):116–125, 1986.

[34] Nikolaos Dimitriou, Anastasios Drosou, and Dimitrios Tzovaras. Scan4reco: towards the digitized conservation of cultural heritage assets via spatiotemporal (4 d) reconstruction and 3 d printing. In *Proceedings of the 14th Eurographics Workshop on Graphics and Cultural Heritage*, pages 53–56. Eurographics Association, 2016.

[35] Julie Dorsey, Holly Rushmeier, and François Sillion. *Digital modeling of material appearance*. Elsevier, 2010.

[36] Mark S Drew, Yacov Hel-Or, Tom Malzbender, and Nasim Hajari. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing*, 30(4-5):317–331, 2012.

[37] Tinsae Dulecha, Andrea Giachetti, Ruggero Pintus, Irina-Mihaela Ciortan, Alberto Jaspe Villanueva, and Enrico Gobbetti. Crack detection in single-and multi-light images of painted surfaces using convolutional neural networks. In *Eurographics Workshop on Graphics and Cultural Heritage*. The Eurographics Association, 2019.

[38] Tinsae G Dulecha, Filippo A Fanni, Federico Ponchio, Fabio Pellacini, and Andrea Giachetti. Neural reflectance transformation imaging. *The Visual Computer*, pages 1–14, 2020.

[39] Nicholas J Durr, Germán González, Daryl Lim, Giovanni Traverso, Norman S Nishioka, Benjamin J Vakoc, and Vicente Parot. System for clinical photometric stereo endoscopy. In *Advanced Biomedical and Clinical Diagnostic Systems XII*, volume 8935, page 89351F. International Society for Optics and Photonics, 2014.

[40] Graeme Earl, Philip Basford, Alexander Bischoff, Alan Bowman, Charles Crowther, Jacob Dahl, Michael Hodgson, Leif Isaksen, Eleni Kotoula, Kirk Martinez, Hembo Pagi, and Kathryn E. Piquette. Reflectance transformation imaging systems for ancient documentary artefacts. In *Proc. International Conference on Electronic Visualisation and the Arts*, pages 147–154, Swindon, UK, 2011. BCS Learning & Development Ltd.

[41] Shireen Y Elhabian, Ham Rara, and Aly A Farag. Towards accurate and efficient representation of image irradiance of convex-Lambertian objects under unknown near lighting. In *Proc. ICCV*, pages 1732–1737. IEEE, 2011.

[42] Hao Fan, Lin Qi, Yakun Ju, Junyu Dong, and Hui Yu. Refractive laser triangulation and photometric stereo in underwater environment. *Optical Engineering*, 56(11):113101, 2017.

[43] Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM TOG*, 26(3):51:1–51:9, 2007.

[44] Peter Fornaro, Andrea Bianco, Aeneas Kaiser, and Lukas Rosenthaler. Enhanced RTI for gloss reproduction. *Electronic Imaging*, 2017(8):66–72, 2017.

[45] Andrew Gardner, Chris Tchou, Tim Hawkins, and Paul Debevec. Linear light source reflectometry. *ACM TOG*, 22(3):749–758, 2003.

[46] Athinodoros S Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *null*, page 816. Ieee, 2003.

[47] Andrea Giachetti, Irina Ciortan, Claudia Daffara, Ruggero Pintus, and Enrico Gobbetti. Multispectral RTI analysis of heterogeneous artworks. In *Proc. GCH*, 2017.

[48] Andrea Giachetti, Irina Mihaela Ciortan, Claudia Daffara, Giacomo Marchioro, Ruggero Pintus, and Enrico Gobbetti. A novel framework for highlight reflectance transformation imaging. *Computer Vision and Image Understanding*, 168:118–131, 2018.

[49] Andrea Giachetti et al. apTool project, 2019. [Online; accessed-March-2019].

[50] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[51] Y Goldman, R Linn, O Shamir, and M Weinstein-Evron. Micro-RTI as a novel technology for the investigation and documentation of archaeological textiles. *Journal of Archaeological Science-Reports*, 19:1–10, 2018.

[52] Hendrik Hameeuw. Mesopotamian clay cones in the ancient near east collections of the royal museums of art and history. *Bulletin van de Koninklijke Musea voor Kunst en Geschiedenis*, 84:5–48, 2015.

[53] Hendrik Hameeuw, Geert Willems, Frank Verbiest, Wim Moreau, Karel Van Lerberghe, and Luc Van Gool. Easy and cost-effective cuneiform digitizing. In *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005)*, pages 73–80. Eurographics Association, 2005.

[54] Øyvind Hammer, Stefan Bengtson, Tom Malzbender, and Dan Gelb. Imaging fossils using reflectance transformation and interactive manipulation of virtual light sources. *Palaeontologia Electronica*, 5(1):1–9, 2002.

[55] Susanna Harris and Kathryn E. Piquette. Reflectance transformation imaging (RTI) for visualizing leather grain surface morphology as an aid to species identification: a pilot study. Archaelogical Leather Group Newsletter 42, 2015.

[56] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[57] Yannick Hold-Geoffroy, Paulo FU Gotardo, and Jean-François Lalonde. Deep photometric stereo on a sunny day. *arXiv preprint arXiv:1803.10850*, 2018.

[58] Zhuo Hui and Aswin C Sankaranarayanan. A dictionary-based approach for estimating shape and spatially-varying reflectance. In *2015 IEEE International Conference on Computational Photography (ICCP)*, pages 1–9. IEEE, 2015.

[59] Zhuo Hui and Aswin C Sankaranarayanan. Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10):2060–2073, 2016.

[60] Satoshi Ikehata. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–18, 2018.

[61] Satoshi Ikehata and Kiyoharu Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2179–2186, 2014.

[62] Satoshi Ikehata, David Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 318–325. IEEE, 2012.

[63] Satoshi Ikehata, David P. Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Photometric stereo using sparse bayesian regression for general diffuse surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(9):1078–1091, 2014.

[64] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. In *CVPR 2011*, pages 2553–2560. IEEE, 2011.

[65] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[66] Eleni Kotoula et al. Reflectance transformation imaging beyond the visible: ultraviolet reflected and ultraviolet induced visible fluorescence. In *Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology, Oxford*, pages 909–919, 2015.

[67] Elini Kotoula and Maria Kyranoudi. Study of ancient greek and roman coins using reflectance transformation imaging. *E-conservation magazine*, 25:74–88, 2013.

[68] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[69] KUL. PLD software KU-Leuven, 2019. [Online; accessed-March-2019].

[70] Anders Landstrom, Matthew J Thurley, and Hakan Jonsson. Sub-millimeter crack detection in casted steel using color photometric stereo. In *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7. IEEE, 2013.

[71] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[72] K LEUVEN. Multispectral microdome, 2015. *URL: https://portablelightdome. wordpress. com/2015/04/29/rich-presents-the-new-multispectral-microdome*, 2.

[73] Yundong Li, Hongguang Li, and Hongren Wang. Pixel-wise crack detection using deep local pattern predictor for robot application. *Sensors*, 18(9):3042, 2018.

[74] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017.

[75] Filippo Lorenzoni, Filippo Casarin, Mauro Caldon, Kleidi Islami, and Claudio Modena. Uncertainty quantification in structural health monitoring: Applications on cultural heritage buildings. *Mechanical Systems and Signal Processing*, 66-67:268–281, January 2016.

[76] Huijuan Lv, Yiheng Cai, and Song Guo. 3D reconstruction of tongue surface based on photometric stereo. In *Proc. Signal Processing (ICSP)*, volume 3, pages 1668–1671. IEEE, 2012.

[77] Lindsay W MacDonald. Colour and directionality in surface reflectance. In *Proc. Conf. on Artificial Intelligence and the Simulation of Behaviour (AISB)*, 2014.

[78] Lindsay W MacDonald. Visual realism in digital heritage. In *Heritage Preservation*, pages 21–45. Springer, 2018.

[79] Lindsay William Macdonald. *Realistic visualisation of cultural heritage objects*. PhD thesis, UCL (University College London), 2015.

[80] Thomas Malzbender, Bennett Wilburn, Dan Gelb, and Bill Ambrisco. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Proc. Rendering techniques*, pages 245–250, 2006.

[81] Tom Malzbender, Dan Gelb, and Hans Wolters. Polynomial texture maps. In *Proc. SIGGRAPH*, pages 519–528, 2001.

[82] Marcello Manfredi, Greg Bearman, Greg Williamson, Dale Kronkright, Eric Doehne, Megan Jacobs, and Emilio Marengo. A new quantitative method for the non-invasive documentation of morphological damage in paintings using rti surface normals. *Sensors*, 14(7):12271–12284, 2014.

[83] Vincent Masselus, Philip Dutré, and Frederik Anrys. The free-form light stage. In *Proc. ACM SIGGRAPH Abstracts and Applications*, pages 262–262. ACM, 2002.

[84] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Trans. Graph.*, 22(3):759–769, July 2003.

[85] Daisuke Miyazaki, Kenji Hara, and Katsushi Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision*, 86(2-3):229, 2010.

[86] Arun Mohan and Sumathi Poobal. Crack detection using image processing: A critical review and analysis. *Alexandria Engineering Journal*, February 2017.

[87] Mark Mudge, Thomas Malzbender, Carla Schroer, and Marlin Lum. New reflection transformation imaging methods for rock art and multiple-viewpoint display. In *VAST*, volume 6, pages 195–202. Citeseer, 2006.

[88] Mark Mudge, Tom Malzbender, Alan Chalmers, Roberto Scopigno, James Davis, Oliver Wang, Prabath Gunawardane, Michael Ashley, Martin Doerr, Alberto Proenca, and Joao Barbosa. Image-based empirical information acquisition, scientific reliability, and long-term digital preservation for the natural sciences and cultural heritage. In *Eurographics (Tutorials)*, 2008.

[89] Mark Mudge, Jean-Pierre Voutaz, Carla Schroer, and Marlin Lum. Reflection transformation imaging and virtual representations of coins from the hospice of the grand st. bernard. In *Proc. VAST*, volume 6, pages 29–40, 2005.

[90] Yasuhiro Mukaigawa, Yasunori Ishii, and Takeshi Shakunaga. Analysis of photometric factors based on photometric linearization. *JOSA A*, 24(10):3326–3334, 2007.

[91] Kate Chapman J. R. Peterson Alistair Cross Mytum, Harold. Reflectance transformation imaging (rti): Capturing gravestone detail via multiple digital images. *Association for Gravestone Studies*, 42(2), 2017.

[92] Sarah E Newman. Applications of reflectance transformation imaging (RTI) to the study of bone surface modifications. *Journal of Archaeological Science*, 53:536–549, 2015.

[93] Fred E Nicodemus. Directional reflectance and emissivity of an opaque surface. *Applied optics*, 4(7):767–775, 1965.

[94] R. Oullette, M. Browne, and K. Hirasawa. Genetic algorithm optimization of a convolutional neural network for autonomous crack detection. In *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, volume 1, pages 516–521 Vol.1, June 2004.

[95] Joseph Padfield, David Saunders, and Tom Malzbender. Polynomial texture mapping: a new tool for examining the surface of paintings. *ICOM Committee for Conservation*, 1:504–510, 2005.

[96] Giampaolo Palma et al. WebRTI Viewer, 2019. [Online; accessed-March-2019].

[97] Gianpaolo Palma, M Baldassarri, MC Favilla, and R Scopigno. Storytelling of a coin collection by means of RTI images: the case of the Simoneschi collection in Palazzo Blu. In *Museums and the Web*, 2014.

[98] Gianpaolo Palma, Massimiliano Corsini, Paolo Cignoni, Roberto Scopigno, and Mark Mudge. Dynamic shading enhancement for reflectance transformation imaging. *Journal on Computing and Cultural Heritage (JOCCH)*, 3(2):1–20, 2010.

[99] J Paluszyński and W Slowko. Surface reconstruction with the photometric method in sem. *Vacuum*, 78(2-4):533–537, 2005.

[100] Rongjiang Pan. Detection of edges from polynomial texture maps. *3D Research*, 7(3):81:1–81:8, 2016.

[101] Thoma Papadhimitri and Paolo Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *International journal of computer vision*, 107(2):139–154, 2014.

[102] Thoma Papadhimitri and Paolo Favaro. Uncalibrated near-light photometric stereo. In *Proc. BMVC*. BMVA Press, 2014.

[103] Vicente Parot, Daryl Lim, Germán González, Giovanni Traverso, Norman S Nishioka, Benjamin J Vakoc, and Nicholas J Durr. Photometric stereo endoscopy. *Journal of biomedical optics*, 18(7):076017, 2013.

[104] Ruggero Pintus, Irina Ciortan, Andrea Giachetti, and Enrico Gobbetti. Practical free-form RTI acquisition with local spot lights. In *Proc. STAG*, 2016.

[105] Ruggero Pintus, Tinsae Dulecha, Alberto Jaspe, Andrea Giachetti, Irina Mihaela Ciortan, and Enrico Gobbetti. Objective and subjective evaluation of virtual relighting from reflectance transformation imaging data. In *GCH*, pages 87–96, 2018.

[106] Ruggero Pintus, Tinsae Gebrechristos Dulecha, I Ciortan, Enrico Gobbetti, and Andrea Giachetti. State-of-the-art in multi-light image collections for surface visualization and analysis. In *Computer Graphics Forum*, volume 38, pages 909–934. Wiley Online Library, 2019.

[107] Ruggero Pintus, Andrea Giachetti, Giovanni Pintore, Enrico Gobbetti, et al. Guided robust matte-model fitting for accelerating multi-light reflectance processing techniques. 2017.

[108] Ruggero Pintus, Kazim Pal, Ying Yang, Tim Weyrich, Enrico Gobbetti, and Holly Rushmeier. A survey of geometric analysis in cultural heritage. In *Computer Graphics Forum*, volume 35, pages 4–31, 2016.

[109] Ruggero Pintus, Simona Podda, and Massimo Vanzi. An automatic alignment procedure for a four-source photometric stereo technique applied to scanning electron microscopy. *IEEE Transactions on Instrumentation and Measurement*, 57(5):989–996, 2008.

[110] Hugo Pires, J Martínez Rubio, and A Elorza Arana. Techniques for revealing 3D hidden archeological features: morphological residual models as virtual-polynomial texture maps. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5):415–421, 2015.

[111] Gilles Pitard, Gaëtan Le Goïc, Hugues Favrelière, Serge Samper, Simon-Frédéric Desage, and Maurice Pillet. Discrete modal decomposition for surface appearance modelling and rendering. In *Optical Measurement Systems for Industrial Inspection IX*, volume 9525, pages 952523:1–952523:10. International Society for Optics and Photonics, 2015.

[112] Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Simon-Frederic Desage, Serge Samper, and Maurice Pillet. Discrete modal decomposition: a new approach for the reflectance modeling and rendering of real surfaces. *Machine Vision and Applications*, 28(5-6):607–621, 2017.

[113] Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Maurice Pillet, Sony George, and Jon Yngve Hardeberg. Reflectance-based surface saliency. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 445–449. IEEE, 2017.

[114] Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Maurice Pillet, Sony George, and Jon Yngve Hardeberg. Robust anomaly detection using reflectance transformation imaging for surface quality inspection. In *Scandinavian Conference on Image Analysis*, pages 550–561. Springer, 2017.

[115] A. Pizurica, L. Platisa, T. Ruzic, B. Cornelis, A. Dooms, M. Martens, H. Dubois, B. Devolder, M. De Mey, and I. Daubechies. Digital Image Processing of The Ghent Altarpiece: Supporting the painting's study and conservation treatment. *IEEE Signal Processing Magazine*, 32(4):112–122, July 2015.

[116] Federico Ponchio, Massimiliano Corsini, and Roberto Scopigno. A compact representation of relightable images for the web. In *Proc. ACM Web3D*, pages 1:1–1:10, 2018.

[117] Federico Ponchio et al. Relight website, 2019. [Online; accessed-March-2019].

[118] Bernhard Preim, Alexandra Baer, Douglas Cunningham, Tobias Isenberg, and Timo Ropinski. A survey of perceptually motivated 3D visualization of medical image data. *Computer Graphics Forum*, 35(3):501–525, 2016.

[119] Eftychios Protopapadakis, Konstantinos Makantasis, George Kopsiaftis, Nikolaos Doulamis, and Angelos Amditis. Crack Identification Via User Feedback, Convolutional Neural Networks and Laser Scanners for Tunnel Infrastructures. In *VISIGRAPP (4: VISAPP)*, pages 725–734, 2016.

[120] Gilles Rainer, Wenzel Jakob, Abhijeet Ghosh, and Tim Weyrich. Neural btf compression and interpolation. In *Computer Graphics Forum*, volume 38, pages 235–244. Wiley Online Library, 2019.

[121] Ramesh Raskar, Kar-Han Tan, Rogerio Feris, Jingyi Yu, and Matthew Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. In *ACM TOG*, volume 23, pages 679–688, 2004.

[122] Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Efstratios Gavves, and Tinne Tuytelaars. Deep reflectance maps. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. to appear.

[123] Peiran Ren, Yue Dong, Stephen Lin, Xin Tong, and Baining Guo. Image based relighting using neural networks. *ACM TOG*, 34(4):111:1–111:12, 2015.

[124] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.

[125] Roland Ruiters and Reinhard Klein. Heightfield and spatially varying brdf reconstruction for materials with interreflections. *Computer Graphics Forum (Proc. of Eurographics)*, 28(2):513–522, April 2009.

[126] Martin Rump, Gero Müller, Ralf Sarlette, Dirk Koch, and Reinhard Klein. Photo-realistic rendering of metallic car paint from image-based measurements. In *Computer Graphics Forum*, volume 27, pages 527–536. Wiley Online Library, 2008.

[127] Tijana Ruzić, Bruno Cornelis, Ljiljana Platisa, Aleksandra Pizurica, Ann Dooms, Wilfried Philips, Maximiliaan Martens, Marc De Mey, and Ingrid Daubechies. Virtual Restoration of the Ghent Altarpiece Using Crack Detection and Inpainting. In *Advanced Concepts for Intelligent Vision Systems*, Lecture Notes in Computer Science, pages 417–428. Springer, Berlin, Heidelberg, August 2011.

[128] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 501–509, 2017.

[129] Imari Sato, Takahiro Okabe, Yoichi Sato, and Katsushi Ikeuchi. Appearance sampling for obtaining a set of basis images for variable illumination. In *Proc. ICCV*, pages 800–807, 2003.

[130] David Selmo, Fraser Sturt, James Miles, Philip Basford, Tom Malzbender, Kirk Martinez, Charlie Thompson, Graeme Earl, and George Bevan. Underwater reflectance transformation imaging: a technology for in situ underwater cultural heritage object-level recording. *Journal of Electronic Imaging*, 26(1):011029, 2017.

[131] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan. Self-calibrating photometric stereo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1118–1125. IEEE, 2010.

[132] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3707–3716, 2016.

[133] Roman Sizyakin, Bruno Cornelis, Laurens Meeus, Maximiliaan Martens, Viacheslav Voronin, and Aleksandra Pizurica. A deep learning approach to crack detection in panel paintings. In *Image Processing for Art Investigation (IP4AI)*, pages 40–42, 2018.

[134] Daniel Smilkov, Nikhil Thorat, Yannick Assogba, Ann Yuan, Nick Kreeger, Ping Yu, Kangyi Zhang, Shanqing Cai, Eric Nielsen, David Soergel, Stan Bileschi, Michael Terry, Charles Nicholson, Sandeep N. Gupta, Sarah Sirajuddin, D. Sculley, Rajat Monga, Greg Corrado, Fernanda B. Viégas, and Martin Wattenberg. Tensorflow.js: Machine learning for the web and beyond. *CoRR*, abs/1901.05350, 2019.

[135] Melvyn L Smith and Richard J Stamp. Automated inspection of textured ceramic tiles. *Computers in Industry*, 43(1):73–82, 2000.

[136] Daniel Soukup and Reinhold Huber-Mörk. Convolutional neural networks for steel surface defect detection from photometric stereo images. In *International Symposium on Visual Computing*, pages 668–677. Springer, 2014.

[137] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[138] Jiuai Sun, Zhao Liu, Yi Ding, and Melvyn Smith. Recovering skin reflectance and geometry for diagnosis of melanoma. In *Computer Vision Techniques for the Diagnosis of Skin Cancer*, pages 243–265. Springer, 2014.

[139] Jiuai Sun, Melvyn Smith, Lyndon Smith, Louise Coutts, Rasha Dabis, Christopher Harland, and Jeffrey Bamber. Reflectance of human skin using colour photometric stereo: with particular application to pigmented lesion analysis. *Skin research and technology*, 14(2):173–179, 2008.

[140] Jiuai Sun, Melvyn Smith, Lyndon Smith, and Abdul Farooq. Sampling light field for photometric stereo. *International Journal of Computer Theory and Engineering*, 5(1):14–18, 2013.

[141] Tatsunori Taniai and Takanori Maehara. Neural inverse rendering for general reflectance photometric stereo. In *International Conference on Machine Learning*, pages 4864–4873, 2018.

[142] Niranjan Thanikachalam, Loïc Baboulaz, Damien Firmenich, Sabine Süsstrunk, and Martin Vetterli. Handheld reflectance acquisition of paintings. *IEEE Transactions on Computational Imaging*, 3(ARTICLE):580–591, 2017.

[143] Christian Tominski, Stefan Gladisch, Ulrike Kister, Raimund Dachselt, and Heidrun Schumann. Interactive lenses for visualization: An extended survey. *Computer Graphics Forum*, 36(6):173–200, 2017.

[144] Hiroyuki Ukida, Tetsuya Sano, Yoshio Tanimoto, and Hideki Yamamoto. 3D shape and color estimation using linear light sources and cameras. In *Proc. Imaging Systems and Techniques (IST)*, pages 1–5. IEEE, 2015.

[145] Marta Díaz-Guardamino Uribe and David W Wheatley. Rock art an digital technologies: the application of reflectance transformation imaging (RTI) and 3D laser scanning to the study of late Bronze Age Iberian Stelae. *Menga: Revista de prehistoria de Andalucía*, (4):187–203, 2013.

[146] Athena Van der Perre, Hendrik Hameeuw, Vanessa Boschloos, Luc Delvaux, Marc Proesmans, Bruno Vandermeulen, Luc Van Gool, and Lieve Watteeuw. Towards a combined use of IR, UV and 3D-imaging for the study of small inscribed and illuminated artefacts. In *Lights On... Cultural Heritage and Museums!*, pages 163–192. FLUP, University of Porto, 2016.

[147] Bruno Vandermeulen, Hendrik Hameeuw, Lieve Watteeuw, Luc Van Gool, and Marc Proesmans. Bridging multi-light & multi-spectral images to study, preserve and disseminate archival documents. In *Proc. Archiving Conference*, volume 2018, pages 64–69. Society for Imaging Science and Technology, 2018.

[148] Vincent Vanweddingen, Chris Vastenhoud, Marc Proesmans, Hendrik Hameeuw, Bruno Vandermeulen, Athena Van der Perre, Frederic Lemmers, Lieve Watteeuw, and Luc Van Gool. A status quaestionis and future solutions for using multi-light reflectance imaging approaches for preserving cultural heritage artifacts. In *Euro-Mediterranean Conference*, pages 204–211. Springer, 2018.

[149] Dylan K Wainwright, George V Lauder, and James C Weaver. Imaging biological surface topography in situ and in vivo. *Methods in Ecology and Evolution*, 8(11):1626–1638, 2017.

[150] Lieve Watteeuw, Bruno Vandermeulen, and Marc Proesmans. On the surface and beyond. an new approach with multispectral photometric stereo to assess illuminated manuscripts and their condition. In *Science and Engineering in Arts, Heritage and Archaeology, book of abstracts*, volume 1, pages 103–103. University College London, 2015.

[151] Michael Weinmann and Reinhard Klein. Advances in geometry and reflectance acquisition (course notes). In *SIGGRAPH Asia 2015 Courses*. ACM, 2015.

[152] Olivia Wiles and Andrew Zisserman. Silnet: Single-and multi-view reconstruction by learning from silhouettes. *arXiv preprint arXiv:1711.07888*, 2017.

[153] Holger Winnemöller, Ankit Mohan, Jack Tumblin, and Bruce Gooch. Light waving: Estimating light positions from photographs alone. In *Computer Graphics Forum*, volume 24, pages 433–438. Wiley Online Library, 2005.

[154] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):513–531, 1980.

[155] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision*, pages 703–717. Springer, 2010.

[156] Tai-Pang Wu and Chi-Keung Tang. Photometric stereo via expectation maximization. *IEEE transactions on pattern analysis and machine intelligence*, 32(3):546–560, 2010.

[157] X-Rite. Total appearance capture ecosystem, 2019. [Online; accessed-March-2019].

[158] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 341–349. Curran Associates, Inc., 2012.

[159] Limin Xie, Zhan Song, Guohua Jiao, Xinhan Huang, and Kui Jia. A practical means for calibrating an led-based photometric stereo system. *Optics and Lasers in Engineering*, 64:42–50, 2015.

[160] Wuyuan Xie, Chengkai Dai, and Charlie CL Wang. Photometric stereo with near point lighting: A solution by mesh deformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4585–4593, 2015.

[161] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J Gortler, David W Jacobs, and Todd Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(37):67–79, 2015.

[162] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)*, 37(4):126, 2018.

[163] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, 2017.

[164] Mingjing Zhang and Mark S Drew. Efficient robust image interpolation and surface properties using polynomial texture mapping. *EURASIP Journal on Image and Video Processing*, 2014(1):25, 2014.

[165] Wenhao Zhang, Mark F Hansen, Melvyn Smith, Lyndon Smith, and Bruce Grieve. Photometric stereo for three-dimensional leaf venation extraction. *Computers in Industry*, 98:56–67, 2018.

[166] Bolun Zheng, Rui Sun, Xiang Tian, and Yaowu Chen. S-net: a scalable convolutional neural network for jpeg compression artifact reduction. *Journal of Electronic Imaging*, 27(4):043037, 2018.

[167] Jinghong Zheng, Zhengguo Li, Susanto Rahardja, Susu Yao, and Wei Yao. Collaborative image processing algorithm for detail refinement and enhancement via multi-light images. In *Proc. ICASSP*, pages 1382–1385. IEEE, 2010.

[168] Simone Zini, Simone Bianco, and Raimondo Schettini. Deep residual autoencoder for quality independent jpeg restoration. *arXiv preprint arXiv:1903.06117*, 2019.

Appendices

APPENDIX A

Supplementary material Neural Reflectance Transformation Imaging

Introduction

In this chapter, we present details quantitative and qualitative experimental results, on both SynthRTI and RealRTI datasets. The complete SynthRTI and RealRTI datasets as well as the NeuralRTI encoding and relighting codes and synthetic RTI rendering tools are accessible from a project webpage `https://univr-rti.github.io/NeuralRTI/`.

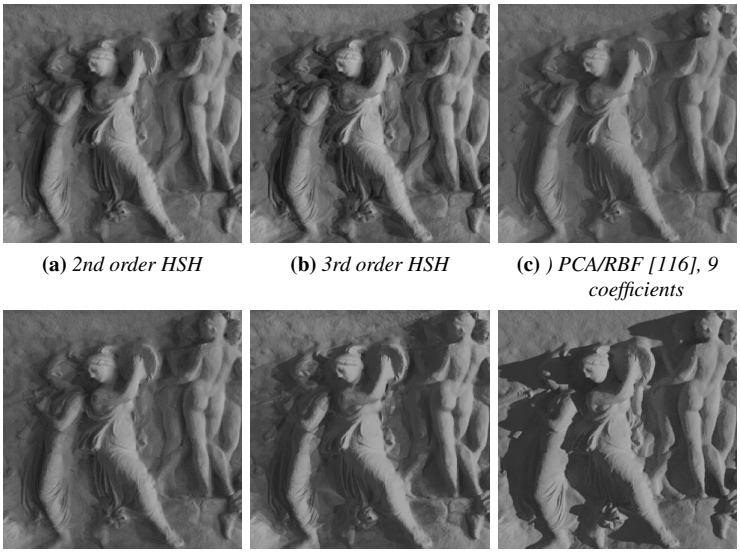
A.0.1 SingleMaterial

Figure A.1 shows relighting results obtained with different methods on the bas-relief object with assigned "matte white" behavior.

Here it is interesting to note that the shadow cast by the different methods. As all methods are local so that no one can deliver a result similar to the ground truth (f) where the low elevation of the light causes large cast shadow. However, it is possible to see that while HSH fitting (a,b) and PCA/RBF interpolation (c,d) result in bright shadow areas and blended artifacts, the result of the Neural RTI relighting (e) presents a darker and more realistic result, even if with borders not corresponding to the real ones, and no blended artifacts.

Figure 5.4 shows relighting results obtained with different methods on the canvas object with assigned "plastic" behavior. Only the Neural relighting (e) can reproduce the specular highlight with reasonable accuracy.

Figures 5.4 and A.2 show that in the case of non-matte materials our novel techniques provide more realistic results both on flat and relieved geometries and bright and dark base colors.



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *) PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

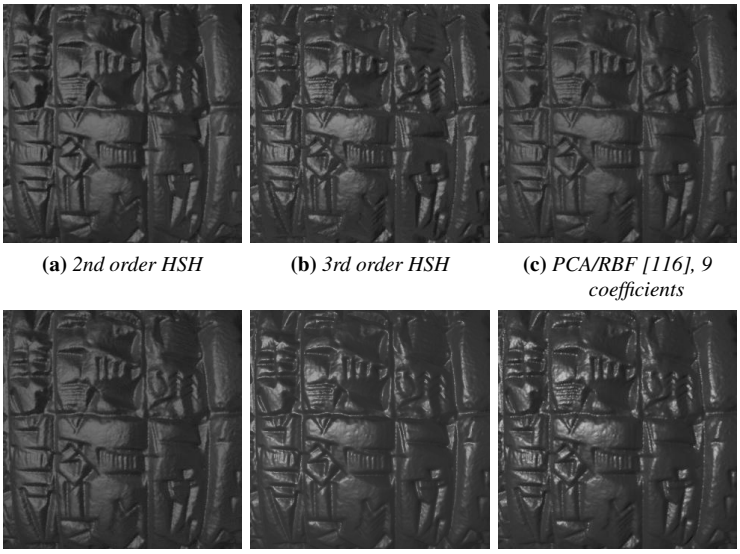
Figure A.1: *Comparison of relighted images from simulated RTI data of the bas relief, "matte" material, light elevation 20 degrees.*

The only material for which NeuralRTI seems to perform worse than the other method is the one with subsurface scattering (Figure A.3). It is interesting to note that, in this case, the rendered images are noisy. Images relighted with NeuralRTI appear noise-free with flattened local variations. The effect may depend also on rendering noise due to the path tracing sampling, but we will investigate it as well as try to generate novel sets with an increased number of realistic materials. Table A.1 shows the average PSNR and SSIM values for the three models with all the materials. Bold fonts indicate the best results. On the bas relief model, creating more shadows, the neural relighting is more often the best method despite the lowest number of parameters.

A.0.2 MultiMaterial

Figures A.4-A.12 show that NeuralRTI is able to better reproduce the high frequency behaviours which are challenging for widely used classical PTM/HSH methods techniques([81], [88]), despite the reduced encoding size. Shadows are definitely darker in A.4. No other methods simulate correctly the specular behavior of the materials in Figure A.6 ,A.7 and A.8, for example. Especially the metallic objects are realistically rendered only by the NeuralRTI method (see Figures A.8-A.12).

Table A.2 shows that NeuralRTI provides the best relighting quality for almost all the material combinations, even compared with the classic 27 and 48 bytes encoding.



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *ground truth image*

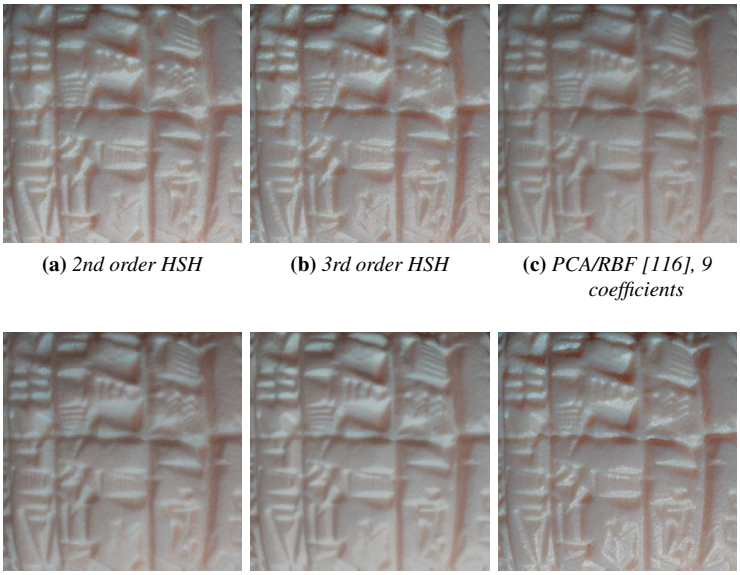
Figure A.2: *Comparison of relighted images from simulated RTI data of a cuneiform tablet, "plastic black" material, light elevation 60 degrees.*

A.1 Relighting results: RealRTI dataset

Finally, we show other examples of relighting results obtained on selected collections from the RealRTI dataset. We still compare the novel technique with classical RTI solutions with equal or higher pixel coding size, this time averaging the results of 5 Leave-One out tests, creating a relightable image on N-1 images of the input dataset and testing quality on the remaining light direction of the input set.

A visual comparison shows that also on the real acquisitions, NeuralRTI is the best option to create plausible specular and shadow components (Figure A.13,A.16). The interpolation of the illumination for the metallic coin is quite realistic and the shadows dark and with sharp contours.

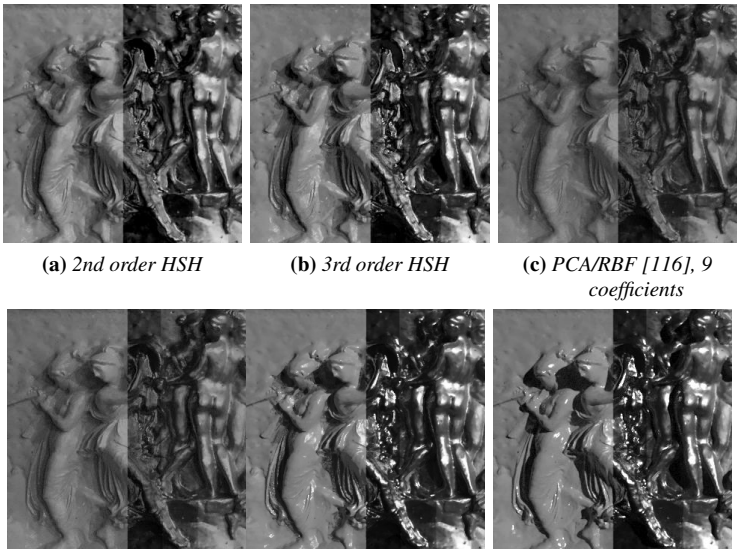
Table A.3 shows that considering the average quality measures, NeuralRTI outperforms the classical RTI techniques on most surfaces. There are only a few cases where other methods like HSH encoded provided better quality scores (even if with 27/48 bytes encoding), e.g. on item 6 (Figure A.15). Typically it happens on flat, matte surfaces, where the low-frequency fitting functions used in classical RTI may help in recovering a more accurate base color. This does not result, however, in the perception of better visual quality.



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.3: *Comparison of relighted images from simulated RTI data of a cuneiform tablet, "subsurface" material, light elevation 40 degrees.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.4: *Comparison of relighted images from simulated RTI data of a bas relief, multimaterial, light elevation 40 degrees.*

	PSNR						SSIM					
Material	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)
1	31.88	46.37	**50.87**	34.88	34.90	41.02	0.994	0.997	**0.997**	0.996	0.996	0.994
2	26.72	30.87	33.78	31.15	31.23	**34.38**	0.982	0.985	**0.988**	0.985	0.987	0.986
3	28.50	36.42	**42.39**	33.22	33.21	35.44	0.991	0.994	**0.995**	0.993	0.993	0.990
4	35.31	40.19	**48.03**	38.55	38.48	41.98	0.993	0.995	**0.997**	0.995	0.995	0.994
5	27.02	27.06	29.46	33.82	34.10	**39.43**	0.954	0.962	0.975	0.979	0.980	**0.987**
6	23.14	26.21	33.15	33.62	34.01	**36.50**	0.953	0.971	0.983	0.980	0.978	**0.986**
7	31.10	40.02	**49.85**	35.09	34.86	36.18	0.992	0.994	**0.995**	0.994	0.994	0.992
8	28.58	36.20	**42.41**	33.27	33.32	35.11	0.987	0.990	**0.991**	0.986	0.989	0.975
average	29.03	35.42	**41.24**	34.20	34.26	37.50	0.981	0.986	**0.990**	0.988	0.989	0.988

(a) Canvas

	PSNR						SSIM					
Material	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)
1	23.02	27.46	**29.19**	26.47	26.77	28.17	0.818	0.845	**0.855**	0.820	0.851	0.859
2	22.09	25.78	**27.60**	23.88	25.09	25.55	0.7716	0.810	**0.8338**	0.765	0.811	0.822
3	22.62	26.84	**28.75**	25.20	26.01	26.33	0.805	0.837	**0.853**	0.797	0.834	0.823
4	29.04	32.47	**34.49**	30.23	31.42	33.11	0.8582	0.887	**0.901**	0.847	0.884	0.892
5	24.16	26.20	28.79	25.29	26.73	**29.09**	0.7484	0.794	0.834	0.764	0.843	**0.852**
6	20.16	23.85	**27.28**	22.31	22.90	25.19	0.737	0.8011	**0.856**	0.768	0.845	0.854
7	24.79	28.58	**30.61**	26.66	27.31	28.54	0.844	0.876	**0.887**	0.834	0.864	0.866
8	24.48	**29.83**	32.61	26.89	28.19	28.50	0.871	0.902	**0.919**	0.814	0.877	0.837
Average	23.79	27.63	**29.92**	25.87	26.80	28.06	0.807	0.844	**0.867**	0.801	0.851	0.850

(b) Cuneiform tablet

	PSNR						SSIM					
Material	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)
1	23.78	26.64	**27.67**	24.94	25.79	25.70	0.811	0.832	**0.836**	0.806	0.836	0.817
2	22.98	25.12	26.41	23.40	24.55	**26.53**	0.764	0.798	0.814	0.764	0.809	**0.834**
3	23.50	26.17	**27.35**	24.39	25.33	25.41	0.797	0.825	**0.834**	0.793	0.828	0.812
4	29.58	32.03	**33.31**	30.19	30.83	32.11	0.877	0.901	**0.907**	0.879	0.901	**0.907**
5	24.81	26.37	28.51	26.39	26.94	**29.16**	0.782	0.818	0.844	0.826	0.867	**0.883**
6	20.56	23.97	**26.80**	22.92	22.98	25.89	0.743	0.814	0.856	0.798	0.849	**0.870**
7	25.10	28.04	**29.19**	26.20	26.67	27.56	0.837	0.872	**0.876**	0.841	0.866	0.872
8	25.46	29.44	**31.30**	25.99	27.38	27.06	0.875	0.900	**0.916**	0.835	0.876	0.849
average	24.47	27.22	**28.82**	25.55	26.31	27.43	0.811	0.845	**0.860**	0.818	0.854	0.855

(c) bas relief, with all the 8 single materials assigned

Table A.1: *Average PSNR and SSIM scores for the relighting from test light directions of the synthetic RTI SingleMaterial data of three basic shapes. Bold fonts indicate best results.*

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

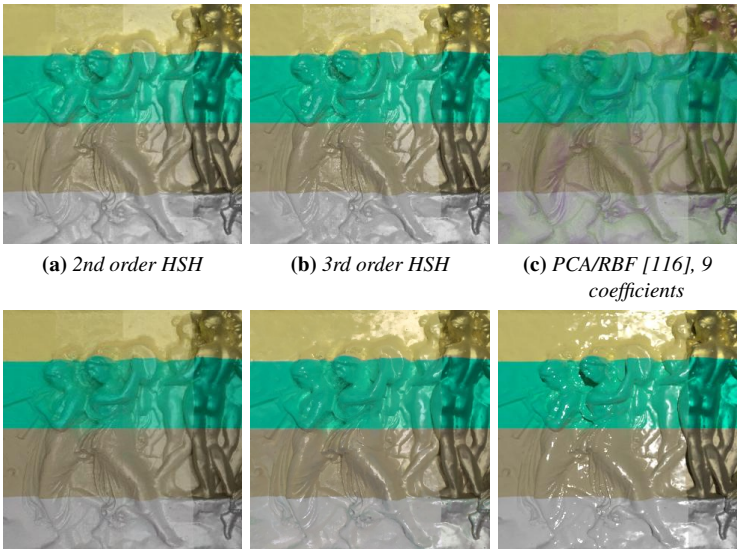
(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.5: *Comparison of relighted images from simulated RTI data of a bas relief, multimaterial, light elevation 20 degrees.*

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.6: *Comparison of relighted images from simulated RTI data of a bas relief, multimaterial, light elevation 60 degrees.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.7: *Comparison of relighted images from simulated RTI data of a bas relief, multimaterial, light elevation 80 degrees.*

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

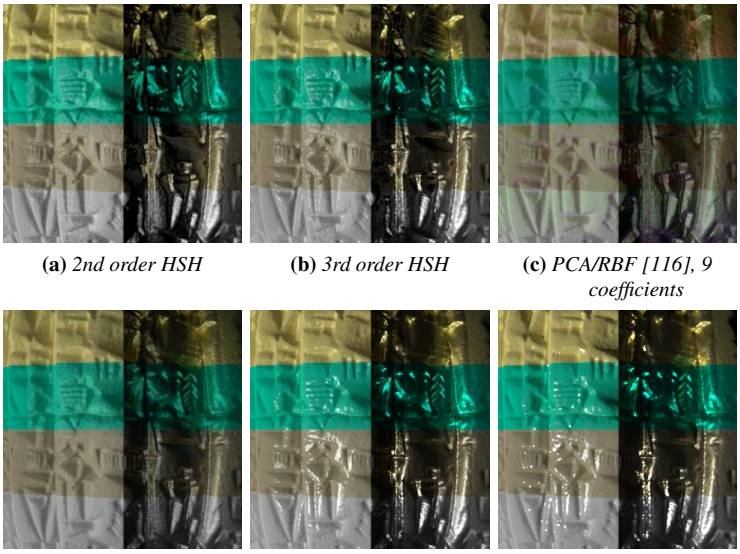
(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.8: *Comparison of relighted images from simulated RTI data of a canvas, multimaterial, light elevation 80 degrees.*

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

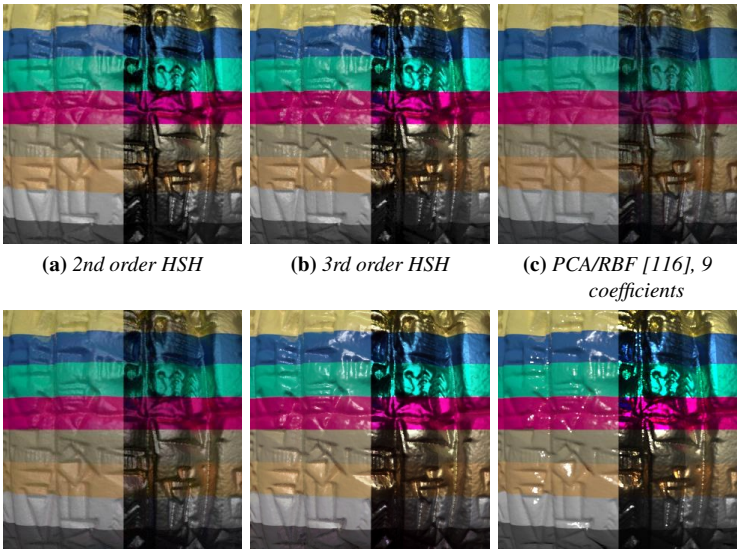
Figure A.9: *Comparison of relighted images from simulated RTI data of a cuneiform tablet, multimaterial, light elevation 20 degrees.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

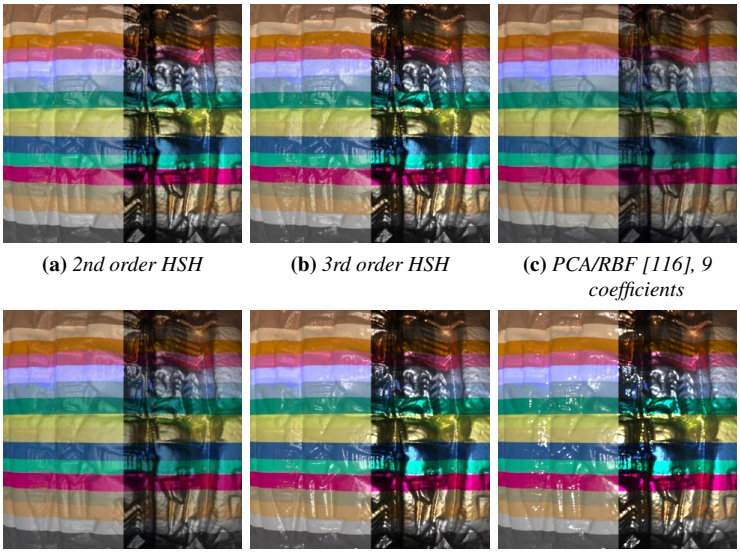
Figure A.10: *Comparison of relighted images from simulated RTI data of a cuneiform tablet, multimaterial, light elevation 40 degrees.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.11: *Comparison of relighted images from simulated RTI data of a cuneiform tablet, multimaterial, light elevation 60 degree.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.12: *Comparison of relighted images from simulated RTI data of a cuneiform tablet, multimaterial, light elevation 80 degrees.*

	PSNR						SSIM					
Material	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)
1	22.99	24.90	26.76	26.14	26.27	**32.60**	0.909	0.884	0.933	0.941	0.943	**0.969**
2	33.03	44.13	**44.83**	35.57	35.59	38.06	0.992	0.996	**0.996**	0.993	0.993	0.989
3	26.51	29.07	31.53	30.97	31.08	**34.42**	0.974	0.987	**0.991**	0.981	0.982	0.986
4	26.07	30.43	31.63	29.97	30.42	**33.51**	0.969	0.978	**0.980**	0.971	0.973	0.976
5	23.77	25.69	26.93	26.34	26.81	**32.32**	0.913	0.893	0.935	0.944	0.947	**0.978**
6	24.06	25.79	27.78	25.89	27.07	**32.01**	0.927	0.913	0.951	0.934	0.948	**0.968**
7	23.03	25.25	27.05	25.62	26.09	**32.58**	0.910	0.888	0.933	0.932	0.939	**0.964**
8	23.67	25.45	26.82	25.69	26.67	**32.29**	0.907	0.884	0.927	0.924	0.939	**0.963**
9	23.39	25.37	26.96	25.33	26.03	**31.58**	0.907	0.887	0.928	0.926	0.934	**0.969**
Average	25.17	28.45	30.03	27.95	28.45	**33.26**	0.934	0.923	0.953	0.949	0.955	**0.973**

(a) Canvas

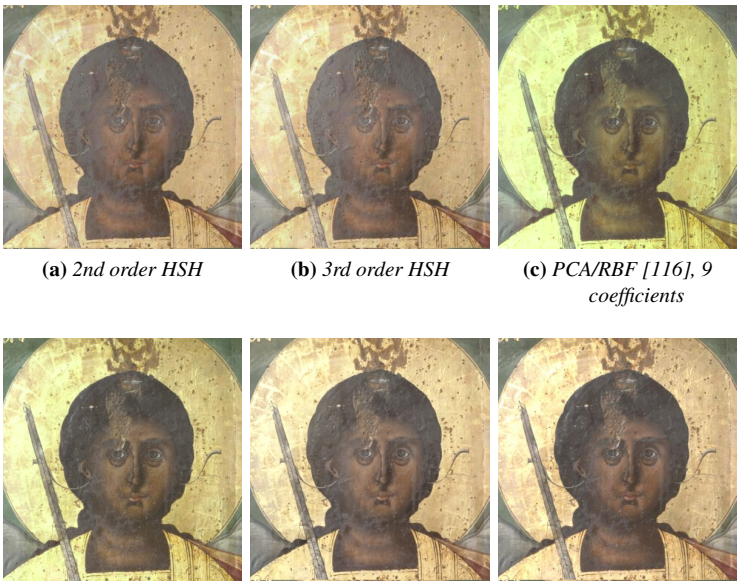
	PSNR						SSIM					
Material	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)
1	19.26	21.25	22.94	20.62	21.35	**23.34**	0.657	0.703	0.734	0.685	0.757	**0.781**
2	23.05	26.35	**27.47**	23.72	25.29	25.71	0.866	0.883	**0.890**	0.831	0.878	0.874
3	21.61	23.99	**25.58**	21.59	23.28	24.05	0.833	0.860	**0.874**	0.798	0.853	0.852
4	21.43	23.50	**24.46**	21.26	22.95	23.40	0.817	0.840	**0.850**	0.765	0.838	0.830
5	20.17	22.16	23.73	20.52	21.74	**23.90**	0.815	0.841	0.864	0.802	0.864	**0.889**
6	20.27	22.30	23.86	20.38	21.81	**23.57**	0.781	0.813	0.836	0.759	0.824	**0.846**
7	19.55	21.63	23.32	19.98	21.33	**23.23**	0.750	0.785	0.813	0.718	0.811	**0.834**
8	19.96	21.89	**23.45**	20.13	21.53	23.36	0.764	0.795	0.820	0.747	0.814	**0.841**
9	19.74	21.74	23.40	19.82	21.28	**23.11**	0.786	0.815	0.840	0.762	0.828	**0.857**
Average	20.56	22.76	24.24	20.89	22.28	**23.74**	0.785	0.815	0.836	0.763	0.830	**0.845**

(b) Cuneiform tablet

	PSNR						SSIM					
Material	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neur.(9)
1	19.35	20.44	21.84	20.53	21.16	**24.34**	0.632	0.654	0.668	0.741	0.772	**0.830**
2	31.51	34.92	**35.62**	25.52	26.15	28.14	0.921	0.935	**0.937**	0.887	0.902	0.911
3	25.38	27.09	**28.81**	22.98	23.98	26.01	0.852	0.873	0.886	0.847	0.871	**0.890**
4	23.76	24.77	25.41	22.51	23.41	**25.17**	0.826	0.838	0.848	0.832	0.864	**0.877**
5	20.59	21.77	23.15	20.85	21.74	**25.14**	0.737	0.757	0.778	0.840	0.868	**0.913**
6	20.58	21.83	23.31	20.72	21.80	**25.27**	0.728	0.749	0.768	0.803	0.840	**0.890**
7	19.63	20.84	22.30	20.17	21.16	**24.50**	0.690	0.715	0.731	0.771	0.827	**0.877**
8	20.24	21.37	22.68	20.37	21.43	**23.69**	0.707	0.727	0.741	0.790	0.830	**0.852**
9	20.04	21.30	22.73	20.26	21.36	**24.27**	0.720	0.739	0.758	0.798	0.840	**0.880**
Average	22.34	23.81	25.10	21.54	22.47	**25.17**	0.757	0.776	0.790	0.812	0.846	**0.880**

(c) bas relief, with all the 9 material mixtures assigned

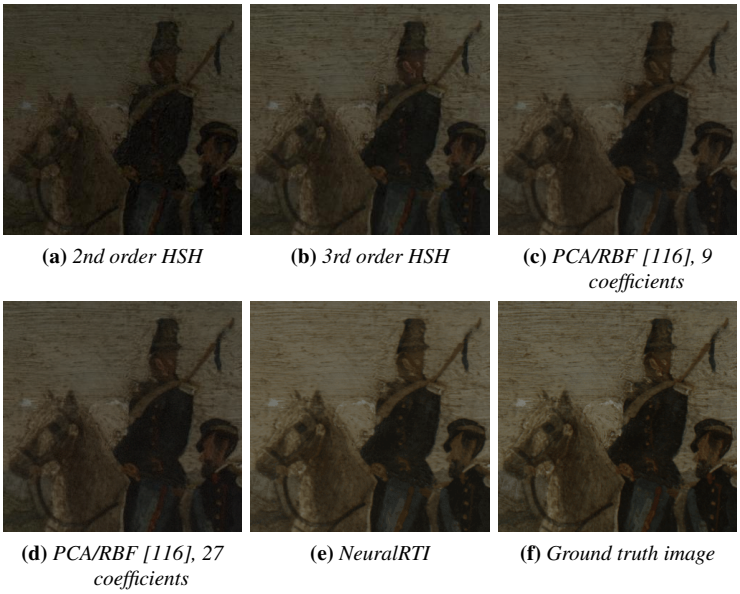
Table A.2: *Average PSNR and SSIM scores for the relighting from test light directions of the synthetic RTI MultiMaterial data of three basic shapes. Bold fonts indicate best results.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

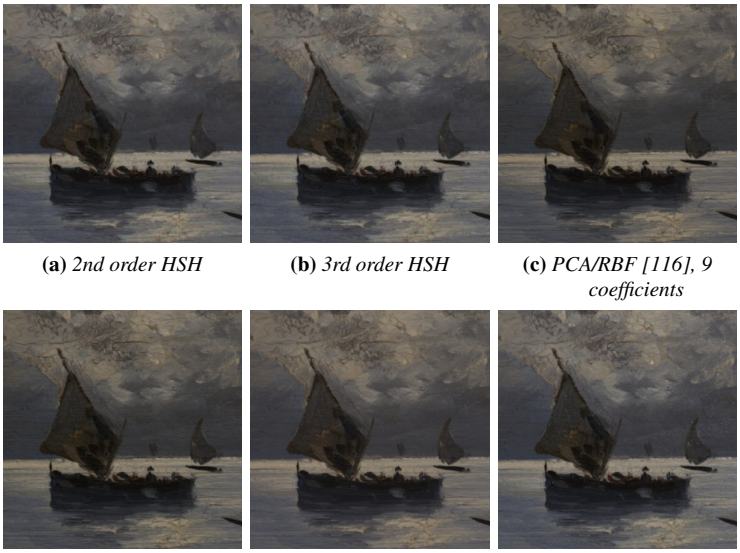
Figure A.13: *Comparison of relighted images from Real RTI data of item3.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.14: *Comparison of relighted images from Real RTI data of item5.*



(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.15: *Comparison of relighted images from Real RTI data of item6.*

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.16: *Comparison of relighted images from Real RTI data of item10.*

(a) *2nd order HSH* **(b)** *3rd order HSH* **(c)** *PCA/RBF [116], 9 coefficients*

(d) *PCA/RBF [116], 27 coefficients* **(e)** *NeuralRTI* **(f)** *Ground truth image*

Figure A.17: *Comparison of relighted images from Real RTI data of item11.*

	PSNR						SSIM					
Item	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neural(9)	PTM(18)	HSH2(27)	HSH3(48)	[116](9)	[116](27)	Neural
1	34.75	39.37	**39.99**	37.06	37.01	38.50	0.981	**0.985**	**0.985**	0.971	0.971	**0.985**
2	28.04	36.28	**38.14**	34.35	34.50	35.53	0.959	0.982	**0.984**	0.973	0.973	0.976
3	24.96	25.14	28.79	28.08	28.62	**31.11**	0.898	0.910	0.921	0.931	0.939	**0.961**
4	26.03	31.02	31.34	32.67	32.92	**34.82**	0.957	0.980	0.981	0.975	0.976	**0.983**
5	31.75	33.22	32.50	25.17	25.39	**37.62**	0.875	0.920	0.896	0.882	0.879	**0.952**
6	33.34	**39.24**	39.91	31.43	31.48	38.43	0.935	0.963	**0.966**	0.898	0.894	0.953
7	32.15	**33.17**	33.81	29.18	28.87	29.07	0.965	**0.975**	0.972	0.957	0.957	0.953
8	29.08	30.68	31.67	29.17	29.18	**31.12**	0.892	0.905	**0.909**	0.874	0.875	0.897
9	21.10	22.34	22.66	21.55	21.82	**25.91**	0.728	0.758	0.742	0.752	0.774	**0.854**
10	16.93	18.64	20.12	18.26	18.59	**22.30**	0.617	0.648	0.656	0.666	0.685	**0.771**
11	29.35	30.10	29.53	27.91	28.25	**31.30**	0.909	0.913	0.886	0.884	0.899	**0.932**
12	29.68	30.24	30.62	28.65	29.17	**31.98**	0.898	0.904	0.883	0.887	0.886	**0.907**
Average	28.10	30.79	31.59	28.62	28.82	**32.31**	0.884	0.904	0.898	0.888	0.892	**0.927**

Table A.3: *Average PSNR and SSIM scores in the Leave-One-Out testing (see main paper) of the relightings performed on the 12 real acquisitions of the RealRTI dataset. Bold fonts indicate the best results.*

A.2 Displaying encoded coefficients

Further interesting analysis can be done by analyzing the meaning of the encoded coefficients. The meaning of the coefficients is not the same for the different RTI dataset, as the decoder is learned, however, a visual analysis of the NeuralRTI encoded byte planes represented as greyscale images can reveal interesting aspects of the captured surface. Figures A.18 and Figures A.19 shows example for two different challenging surfaces. It is possible to see that some coefficients seem related to specific material properties (roughness, glossiness).

Figure A.18: *Image sample (top left) and the nine NeuralRTI encoded byte planes represented as greyscale images for a coin RTI encoding*

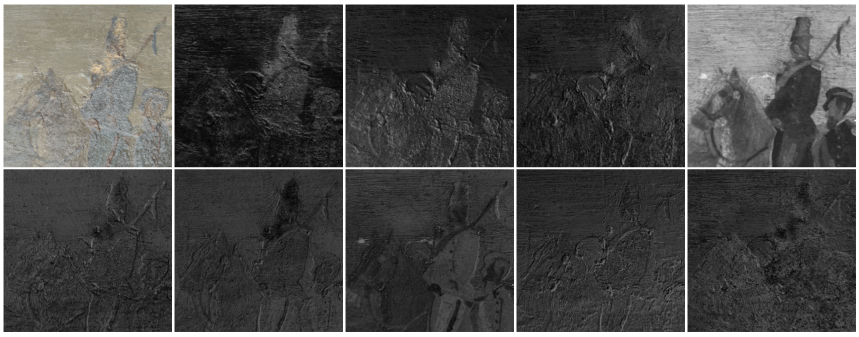


Figure A.19: *Image sample (top left) and the nine NeuralRTI encoded byte planes represented as greyscale images for a painting RTI encoding.*