TU WIEN Informatics

# Multi-Light Imaging for Graphical Heritage

DISSERTATION

zur Erlangung des akademischen Grades

**Doktor der Technischen Wissenschaften**

eingereicht von

**Simon Brenner**
Matrikelnummer 0927175

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

Diese Dissertation haben begutachtet:

Roger L. Easton

José Manuel Menéndez

Wien, 15. Jänner 2024

Simon Brenner

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Informatics

# Multi-Light Imaging for Graphical Heritage

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

**Doktor der Technischen Wissenschaften**

by

**Simon Brenner**
Registration Number 0927175

to the Faculty of Informatics

at the TU Wien

Advisor: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

The dissertation has been reviewed by:

Roger L. Easton

José Manuel Menéndez

Vienna, 15th January, 2024

Simon Brenner

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

# Erklärung zur Verfassung der Arbeit

Simon Brenner

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 15. Jänner 2024

Simon Brenner

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

# Danksagung

Das Abhängige Entstehen ist ein zentrales Konzept im Buddhismus - es bedeutet, dass kein Phänomen aus sich selbst heraus existiert, sondern nur aufgrund von Bedingungen und Verbindungen. Ähnlich verhält es sich mit der vorliegenden Dissertation, die nur durch das Zutun von unzähligen Menschen entstehen konnte.

Als erstes möchte ich mich bei meinem Betreuer Robert und allen anderen großartigen Kolleginnen und Kollegen am Computer Vision Lab bedanken, die für sieben äußerst angenehme und trotzdem produktive Arbeitsjahre verantwortlich sind und damit am Übertauchen gelegentlicher Sinnkrisen maßgeblich beteiligt waren.

Außerdem bedanke ich mich bei den vielen Kolleginnen und Kollegen anderer Institutionen, mit denen ich zusammenarbeiten durfte. Sie haben einen Großteil meiner Arbeit erst ermöglicht - sei es durch das Aufwerfen interessanter Fragestellungen, den Zugang zu historischen Objekten oder, nicht zuletzt, das gemeinsame Aufstellen von Finanzierungen.

In dem Zusammenhang gilt mein Dank auch dem österreichischen Volk, das mit seinem Steuergeld sowohl meine Ausbildung als auch meine Forschungsarbeit (und damit meinen Lebensunterhalt) finanziert hat. Ich hoffe, dass sich diese Investition irgendwann lohnen wird.

Schließlich möchte ich meiner Familie danken, die mich schon immer unterstützt hat; meiner Frau Manu, für das glückliche Leben das sie mir bereitet (sowie das gelegentliche Soda Zitron zur Aufmunterung beim Schreiben); und meinen Töchtern Nora und Lina, die mit ihrem Geburtstermin die notwendige Deadline für den Abschluss dieser Arbeit gesetzt haben.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

# Acknowledgements

Dependent Origination is a central concept in Buddhism - it means that no phenomenon exists out of itself, but only because of conditions and connections. Similarly, this dissertation could only come into being through the contributions of countless people.

First of all, I would like to thank my supervisor Robert and all the other great colleagues at the Computer Vision Lab, who have been responsible for seven extremely pleasant yet productive years of work, and have thus been instrumental in overcoming occasional crises of purpose.

I would also like to thank the many colleagues from other institutions with whom I have had the privilege of working. They have made much of my work possible - whether by raising interesting questions, providing access to historical objects, or, last but not least, jointly acquiring funding.

In this context, I would also like to thank the Austrian people, whose tax money has financed both my education and my research (and thus my livelihood). I hope that this investment will pay off someday.

Finally, I would like to thank my family, who have always supported me; my wife Manu, for the happy life she is giving me (as well as the occasional soda lemon for motivation while writing); and my daughters Nora and Lina, whose birth date set the necessary deadline to complete this work.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

# Kurzfassung

Historische Artefakte der Kategorie Graphical Heritage zeichnen sich durch lokale Veränderungen von physischen Oberflächen aus, um textliche oder bildliche Inhalte zu vermitteln - z.B. Manuskripte, Zeichnungen, Gravuren oder Reliefs. Wenn sie durch natürlichen Verfall oder absichtliche Veränderungen beeinträchtigt sind, können die Inhalte mit bloßem Auge oder herkömmlichen Digitalisierungsmethoden unzugänglich sein. Multi-Light Imaging beschreibt Methoden, bei welchen ein Objekt unter verschiedenen Beleuchtungsbedingungen erfasst wird – das beinhaltet die Multispektralfotografie (Variation der Lichtspektren) und Photometric Stereo (Variation der Lichtrichtungen). Die Multispektralfotografie wird zur Erfassung chemischer Oberflächenvariationen durch Tinten und Pigments verwendet, während Photometric Stereo Variationen der Oberflächengeometrie erfasst.

Mit Low-Level-Operationen wie Bilderfassung, Kalibrierung, Oberflächenrekonstruktion und Visualisierung geht die Arbeit auf die Bedürfnisse von Geisteswissenschaftlern ein, die qualitativ hochwertige, interpretierbare Bilder für ihre Untersuchungen benötigen: Ein mobiles Bildaufnahmesystem wird beschrieben, ergänzt durch Ansätze zur Messung und Kompensation von longitudinalen chromatischen Aberrationen, die bei der Multispektralfotografie auftreten, sowie Heuristiken für die Nachbearbeitung und Kalibrierung. Allgemeine Beiträge zu Photometric Stereo mit besonderer Anwendbarkeit auf Graphical Heritage werden vorgestellt: Die Rekonstruktion von annähernd flachen Objekten mit unterschiedlicher Anzahl und Anordnung von Lichtquellen wird evaluiert; dabei wird festgestellt, dass mit einer kleinen Anzahl von kreisförmig angeordneten Lichtquellen kuppelförmige Anordnungen mit vernachlässigbarem Qualitätsverlust ersetzen werden können. Weiters werden Fehler, die durch ein vereinfachtes Beleuchtungsmodell verursacht werden, theoretisch und empirisch untersucht; basierend darauf wird eine Strategie zur Fehlerminimierung formuliert, die sich besonders für die Aufnahme von Graphical Heritage (auf annähernd flachen Oberfläche) unter räumlichen Einschränkungen (z.B. mit einem tragbaren System in einer Bibliothek) eignet. Eine Fallstudie, in der Gedichte aus Vertiefungen in Papier rekonstruiert werden, demonstriert die Anwendung der vorgeschlagenen Techniken.

Ein weiterer Teil der Arbeit befasst sich mit der Entwicklung von Methoden zur Bewertung der Qualität von Bildmaterial - insbesondere mit der objektiven Bewertung der Lesbarkeit von Text in Bildern. Zu diesem Zweck wird ein neuer Datensatz von Manuskriptbildern vorgestellt, die von Geisteswissenschaftlern hinsichtlich ihrer Lesbarkeit bewertet wurden.

Aufgrund der Neuartigkeit des Studiendesigns werden auch die Kohärenz und Validität der gewonnenen Bewertungen durch verschiedene statistische Analysen nachgewiesen. Auf der Grundlage dieses Datensatzes wird ein Rahmen für die Prüfung potenzieller quantitativer Schätzer für die Lesbarkeit von Text in Bildern definiert und eine Reihe von potenziellen Methoden getestet, um eine Grundlage für weitere Forschung zu schaffen.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

# Abstract

Historical artifacts in the category of graphical heritage are characterized by localized modifications of physical surfaces to convey textual or pictorial contents, such as manuscripts, drawings, engravings or reliefs. When affected by natural degradation or purposeful alterations, contents may be inaccessible with the naked eye or conventional digitization methods. Multi-light imaging involves capturing an object under various lighting conditions, encompassing multispectral imaging (variation in light spectra) and photometric stereo (variation in light directions). Multispectral imaging is used to record chemical surface variations resulting from inks and pigments, while photometric stereo captures variations in surface geometry and depth.

By focusing on low-level operations such as image acquisition, calibration, shape reconstruction and visualization, the thesis caters to the needs of humanist scholars who require high-quality interpretable imagery for their investigations: A mobile image acquisition system is described, complemented by approaches for measuring and compensating longitudinal chromatic aberrations occurring in multispectral imaging, and heuristics for post-processing and calibration. General contributions on photometric stereo with particular applicability to graphical heritage acquisition are made: Photometric stereo reconstruction with varying numbers and arrangements of light sources is evaluated, finding that for mostly flat surfaces, a small number of circularly arranged light sources can replace dome-shaped setups with negligible quality loss; a theoretical and empirical analysis of the errors introduced by a simplified lighting model gives rise to a general error mitigation strategy, especially useful for recording graphical heritage (on a mostly flat surface) under spatial constraints (*e.g.*, with a portable system in a library). A case study in which poems are retrieved from indentations in paper demonstrates the application of the proposed techniques.

Another line of work is concerned with developing methods for evaluating the quality of graphical heritage imagery - in particular, with objectively assessing human text legibility in images. To this end, a novel dataset of manuscript images, rated for legibility by humanist scholars, is introduced. Being created with a novel study design, the coherence and validity of expert ratings obtained is shown with several statistical analyses. Based on this dataset, a framework for testing potential quantitative estimators for text legibility in images is defined and a set of candidate methods is tested in order to create a baseline for further research.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

# Contents

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

# CHAPTER 1

# Introduction

What's this? What's this?
There's color everywhere.

*Jack Skellington*

The title of this thesis - "Multi-Light Imaging for Graphical Heritage" - attempts to concisely summarize the range of topics that are treated in the work. To be specific, definitions of the constitutive terms are given in the following paragraphs:

**Graphical heritage** is used as an umbrella term for historical artifacts characterized by local manipulations of physical surfaces for conveying textual or pictorial contents. This includes artifacts created by the application of inks and pigments, such as manuscripts and drawings, but also patterns created by local modifications of shape, such as engravings, chasings and reliefs. In this sense, graphical heritage means the information intended to be conveyed rather than the physical object used as a medium.

**Multi-light imaging** refers to methods that rely on imaging an object multiple times under the same viewing conditions but under varying lighting conditions. In this work, multispectral imaging (where light spectra are varied) and photometric stereo (where light directions are varied) are considered. While the former is applied to record chemical surface variations (due to the application of inks and pigments), the latter is sensitive to variations in surface geometry and depth.

Following these definitions, multi-light imaging methods are predestined for recording graphical heritage. As widely demonstrated in literature (see Chapter 2), their potential for capturing more of the physical aspects of an object than conventional digitization makes them valuable for comprehensive documentation, but above all, as a data source for the recovery of contents that are contained in the material but hard to access with the naked eye. Considering the various levels of abstraction associated with computer and machine vision, from measuring radiation intensity to the extraction of semantic information [Dav12], this thesis is mainly concerned with low-level operations, such as image acquisition, calibration, shape reconstruction and image transformations for visualization; semantic analysis is consciously left to the human(ist) target audience. This orientation is conditioned by the multi-disciplinary work in the context of which the contributions of this thesis were generated: dealing with small numbers of unique and heterogeneous objects, the cooperating scholars above all demanded high-quality interpretable imagery as an additional data source for their investigations. This immediately leads to a problem that is one of the major research topics of this thesis: if methods aim at producing images that provide information about previously inaccessible contents to a human viewer, how can the quality of the results be assessed objectively?

The following section elaborates on the challenges in multi-light imaging for graphical heritage that motivate the contributions of this work. At the end of this chapter, these contributions are explicated and the structure of the thesis is outlined.

## 1.1 Motivation

Accessing the textual and pictorial contents left by our predecessors can be challenging for various reasons: from the degradation of substrate materials and colorants due to suboptimal storage conditions or disasters [WCCM00, TSA$^+$19], up to purposeful deletions for both pragmatic and ideological reasons [CG07, pp.108-111]. As typically the authors of such contents cannot be consulted, scholars must base their reconstructions on evidence available today [Joh19, pp.13-15]. This work is dedicated to enhancing the accessibility of graphical heritage preserved in material remains, thus increasing the available evidence.

Graphical heritage is divided into two fundamental classes, given by their means of creation; they also define the imaging methods appropriate for their accessing and documentation (whereby the occurrence of both classes on a single object is not unusual); Table 1.1 gives an overview, and examples are presented in the following.

| | *manifestation* | *means of creation* | *means of documentation* |
|---|---|---|---|
| **1.** | **reflectance variations** | application of inks, pigments or dyes | multispectral imaging |
| **2.** | **depth variations** | engraving, chasing, embossing, carving | photometric stereo |

Table 1.1: Two classes of graphical heritage

### 1.1.1 Challenging graphical heritage

The types of graphical heritage described below are real-world cases, in which conventional photography or direct visual examination were not sufficient to document or access the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

contents. The list is certainly not exhaustive but gives an impression of challenges faced in scholarly practice.

**Medieval manuscripts** are predominantly written on parchment, which is made from animal skin - mainly calf, sheep and goat - via a lengthy preparation process in which the hides are cleaned of hair and fat, stretched and dried under tension [CG07, pp.9-13]. Depending on the purpose, the resulting leaves would be used individually or combined into codices (books) or scrolls [CG07, pp.49-53, 250]. Texts are predominantly written with iron gall ink made from iron salts and tannins from oak galls, resulting in a deep black color [CG07, p.19]. Various pigments were used for decorated initials, miniature paintings and other ornaments. As parchment was an expensive resource, it was frequently recycled: text not deemed useful was washed or scraped off and overwritten with newer text, resulting in so-called palimpsests [CG07, pp.108-110]. Accessing the earlier texts of these palimpsests is a prominent application scenario for multispectral imaging [Kög14, EKC03, HGS12]; Figure 1.1a shows an example. Other reasons rendering manuscript contents inaccessible are degradation due to environmental influences [CBB03], or covering with other materials [RBC+19]; examples of recovering such contents with multispectral imaging and infrared photography are shown in Figures 1.1b and 1.1c respectively. Aside from these contents that fall in the category of reflectance variations, medieval manuscripts also contain depth variations that are of interest for scholarly analysis, for example, dry-point ruling lines and pricking or dry-point glosses [CG07, pp.16-17,45]. Figure 1.1d shows a manuscript that contains both ruling and glosses in dry-point; in the photometric stereo based visualization, the dry-point glosses (bottom right) are darker than the background, the ruling lines made from the opposite side are brighter than the background, and the main text is overlaid in blue.

**Modern Documents** can pose similar challenges as medieval manuscripts, although the production techniques and materials involved differ. Figure 1.1e shows a business document stained with printing ink, where the handwriting is separated from the staining via multispectral imaging (this examination aimed at determining the origin of the document, as the reverse side of the document contains a drawing suspected to be an unknown preliminary sketch of a famous painting). Figure 1.1f shows a letter of poet W. A. Auden, on which inkless typewriter impressions were discovered. Only with visualizations through photometric stereo, the imprinted texts could be read and identified as an unpublished intermediate version of a poem [BFM23].

**Etruscan Mirrors** are among the most typical objects remaining of the Etruscan period [DG85]. Exemplars of the grip mirror type are in essence a bronze disc of which one side was polished to serve as a mirror, while the other side was frequently decorated with engraved/chased line drawings [DG85]. Unless these drawings are visually enhanced by applying a white substance [Nol29],[CB13, p.88] (a questionable practice by modern conservational standards), documentation can be challenging due to their subtlety and the presence of corrosion. The engravings shown in Figure 1.1g are partially treated

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

with the white substance, making them visible in a conventional photograph; in the photometric stereo reconstruction (high-pass filtered depth map), the entire drawing is visible. The example shown in Figure 1.1g is heavily corroded; nevertheless, the drawings are still present in the affected parts and were recorded with photometric stereo (the visualization shows the normal map).

After giving a taste of the applications of multi-light imaging for graphical heritage, the remainder of this section discusses the specific challenges that are addressed in this work, ranging from estimating and correcting errors during image acquisition to the problems of evaluating the results of imaging or image processing methods with respect to their usefulness for accessing graphical heritage.

### 1.1.2 Aims of this work

Viewed at a conceptual level, this work aims at recording and visualizing surface variations of physical objects, such that the results *serve as evidence* for the discovery and characterization of graphical heritage in humanist research. This translates to the following specific research aims:

1. Optimizing multi-light imaging methods and associated low-level processing for the acquisition and visualization of graphical heritage.
2. Developing objective evaluation methods for the quality of images visualizing graphical heritage.

With regard to the first aim, the goals of the optimizations must be further specified: Requirements on imaging processes are imposed by the nature of objects and their institutional and physical embedding; requirements on the resulting images are imposed by the viewers and their applications. Typically the objects are not easily transportable (for physical or logistical reasons), such that the imaging must be done on-site. This leads to constraints on the amount and size of equipment that can be used, the conditions under which imaging is performed, and the time frame in which the acquisitions must be completed. Thus, one aim of the work is to design and test a system for Multi Light image acquisition that is transportable, flexible with respect to work environments, and enables an efficient and safe imaging workflow.

Of the requirements for the imagery produced, some are straightforward and easy to assess: for detailed inspection by human researchers, the images should have a spatial resolution sufficient to depict the smallest detail of interest; this also encompasses sharpness, as blurring attenuates high frequencies and thus reduces effective resolution [MCR18]. Furthermore, images should record properties of the imaged surface - for example, depth variations reconstructed from photometric stereo should correspond to depth variations of the surface, and intensity values recorded with multispectral imaging should be tied to reflectance variations of the surface. Another aim is thus to optimize imaging and

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

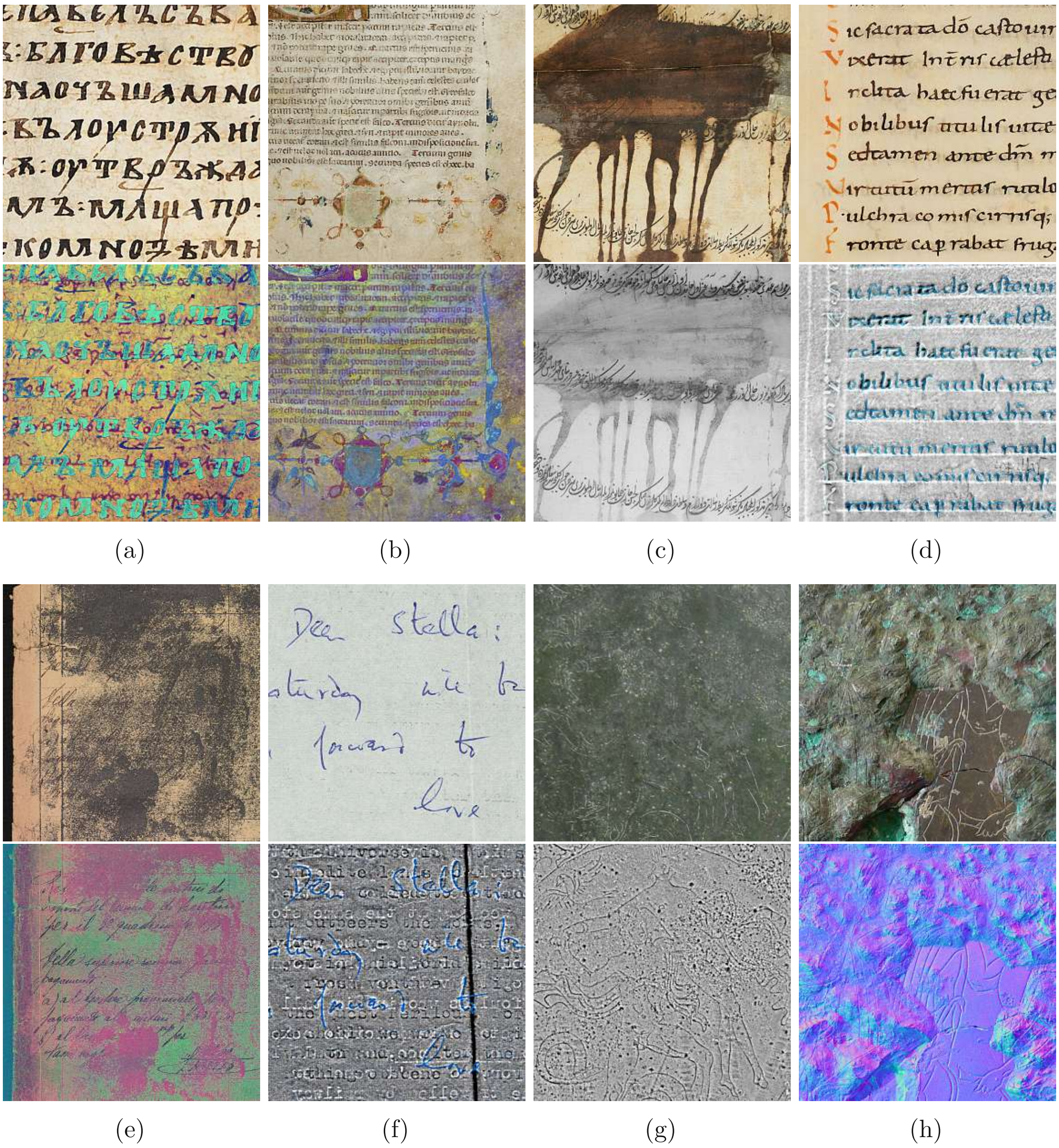


Figure 1.1: Examples for graphical heritage visualized by multi-light imaging methods: (a) the undertext of a palimpsest shows in a false color image from independent components of a multispectral image. (b) the shapes of degraded ornaments are visualized in a false color image based on principal components of a multispectral image. (c) an ottoman charter covered with spilled ink is recovered in an infrared image. (d) dry-point glosses and ruling are visualized in a composite image of high-pass filtered depth map and albedo map obtained via photometric stereo. (e) handwriting and staining from printing ink is separated with multispectral imaging and principal component analysis. (f) typewriter impressions on a letter are visualized with photometric stereo, where the darker text is impressed from the viewer's side and the brighter text from the opposite side; see Section 4.4 for details. (g,h) engravings on an Etruscan mirror visualized with photometric stereo, via depth map and normal map respectively. All visualizations are based on imagery acquired with the system described in Section 3.1.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

image processing with respect to these requirements, while respecting the constraints of the last paragraph. Specifically, the problems of focus shifts in multispectral imaging and ways of dealing with suboptimal lighting configurations in photometric stereo are addressed. Most importantly, however, the images produced should show the graphical heritage contained in the surface as completely and distinctly as possible. The evaluation of this image property, however, is not trivial. Thus, one line of work associated with the second research aim is dedicated to developing appropriate evaluation metrics. It must be noted here, however, that said evaluation metrics are not used to evaluate any other contributions of this work - first, because at the time of writing, they are not mature enough and second, evaluating own work with a self-defined metric would be dubious.

To demarcate the domain of this work, it is important to stress that discovery and characterization themselves are not in the domain of computer vision. When sticking to the definition of graphical heritage as the contents created by our predecessors, making statements about it equals to making statements about the past: what was written/painted/engraved on a particular object by its creator? Making statements about human processes of the past, however, falls in the domain of humanist disciplines like archaeology and history [Joh19, p.15]. These disciplines are specialized in developing plausible theories about past events, grounded on pieces of evidence existing in the present [Joh19, pp.13-15,41]; in this context, multi-light imaging can augment the available evidence by visualizing properties of objects that are inaccessible or unnoticed otherwise [Arn14]. It is then up to the investigating scholar to decide, incorporating contextual information about the object at hand, if the intensity variations in the resulting images are likely to result from actual remains of graphical heritage, or from other sources such as variations in the carrier material, dirt, or artifacts caused by imaging or processing.

Computer vision methods aiming to extract higher level information from images, such as transcription of text or classification of depictions, are relevant for applications where large numbers of comparable objects must be processed and *mostly correct* results are acceptable - a striking example is the *Transkribus* handwritten text recognition platform, which is used to make digitized archives of handwritten documents searchable [MST+19]. For the applications that have inspired this thesis, however, such attempts are both impractical and irrelevant, as they are concerned with small numbers of heterogeneous objects with *a priori* unknown contents, often manifesting only in subtle details - see Figure 1.1 for examples.

## 1.2 Contributions

A large part of the contributions of this work are concerned with the improvement of image acquisition methods and low-level processing for graphical heritage documentation, with both multispectral imaging and photometric stereo; another line of work addresses the evaluation of imagery with respect to its usefulness for human viewers. The individual contributions are summarized as follows:

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

- A mobile image acquisition system for multispectral imaging and photometric stereo is described. Originally developed for the imaging of historical manuscripts, this system is used for all subsequent experiments and use cases.

- As additions to the imaging system, approaches for measuring and compensating longitudinal chromatic aberrations ('focus shifts') occurring in multispectral imaging and for light source calibration for photometric stereo are described.

- The quality of photometric stereo reconstruction with varying numbers and arrangements of light sources is evaluated, finding that for mostly flat objects, a dome-shaped acquisition system can be reduced to a single ring of light sources located at an optimal elevation angle.

- A theoretical and empirical analysis of the errors in photometric stereo introduced by a simplified lighting model (the parallel light assumption) is conducted.

- Grounded in the findings of the previous point, a general error mitigation strategy is formulated. It is especially useful for recording graphical heritage (where the base surface tends to be approximately flat) under spatial constraints (such as with a portable light dome in a library).

- The contributions to photometric stereo described above are combined and demonstrated in a case-study, where previously unknown versions of poems by W. A. Auden were retrieved from indentations in paper.

- The use of artificially degraded manuscripts for assessing the quality of image-based text restoration methods is assessed using a public dataset.

- A novel dataset of manuscript images, rated for text legibility by humanist scholars, is introduced. Being created with a novel study design, the coherence and validity of expert ratings obtained are shown with several statistical analyses.

- Based on this dataset, a framework for testing potential quantitative estimators for text legibility in images is defined; with it, a set of candidate methods is tested in order to create a baseline for further research.

A list of peer-reviewed publications intersecting with these contributions is found in the Appendix.

## 1.3 Research paradigms and methodologies

Three prevailing research paradigms or traditions are identified in computer science, that differ in their ontological (what are the subjects of study?) and epistemological (what can we know about them?) assumptions [Weg76, Ede07, TS08]. While an in-depth discussion of the application areas, merits and flaws of each of these paradigms would go beyond

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

the scope of this work (and has been held elsewhere [Weg76, Ede07, TS08]), a concise summary is attempted in the following:

- The **mathematical** (or formalist/rationalist) tradition views computer programs as mathematical objects and relies on deductive reasoning in order to prove their correctness; like in mathematics, the thus gained knowledge is *a priori*, *i.e.*, testing is not necessary [Ede07, TS08].
- The **engineering** (or technological/technocratic) tradition views programs as artifacts created by people, developed to perform certain tasks (by arbitrary means). Knowledge about them is gained by testing and is thus *a posteriori* and subject to uncertainty; *i.e.*, testing a program cannot prove its correctness [Ede07, TS08]. In general, engineering aims at understanding how to build useful things [Dav98, p.8].
- The **scientific** (or empirical) tradition views computer programs as information processes occurring in the world, be it naturally or artificially. Knowledge is gained through "the scientific method", commonly characterized by the postulation and testing of hypotheses about the studied phenomena [Ede07, TS08] (although both the existence and adequacy of a universal "scientific method" in scientific practice are highly debatable [Kuh62, Fey75]). A distinguishing feature from the engineering tradition is the degree of abstraction and claim for generality of the sought knowledge [TS08].

Regarding this thesis, the aims and methods featured are difficult to attribute to one single paradigm, and are perhaps best described as a mixture of engineering and scientific traditions. An assessment of the individual contributions with respect to their underlying paradigms is attempted in the following.

The contributions discussed in Chapter 3 are concerned with practical aspects of image acquisition, from hardware setups to calibration methods. They are clearly conducted in the spirit of engineering, with the main objective being the development of useful things and to solve problems [TS08]. The general methodology consists in constructing solutions - using mathematics, scientific results, heuristics and intuition with equal merit - and testing their performance. The *a posteriori* knowledge gained this way is inherently uncertain and its generalizability is limited [TS08].

Chapter 4, where aspects of photometric stereo are investigated, leans more towards a scientific paradigm: the reconstruction of surface normals from input images is viewed as an information process as a phenomenon of the world, which can be studied in terms of observation, formulation of hypothesis, and controlled experiments [TS08]; the expected outcome is general knowledge about photometric stereo reconstruction processes. For example in Section 4.3, a hypothesis on the dependence of reconstruction quality on object size and light source distance is formulated - based on observations and theoretical considerations - and tested in experiments. Nevertheless, also elements of engineering are found in the contributions; generally, in motivating their relevance with usefulness

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

for designing more efficient acquisition setups, and specifically, in the error mitigation strategy described in Section 4.3.1. Also the case study presented at the end of the chapter is in essence the demonstration of the usefulness of previously described concepts, complemented with backgrounds from literary studies - and is thus situated in the engineering tradition.

A similar situation is encountered in the works on image quality assessment presented in Chapter 5. On one hand, the contributions are driven by the goal of developing a method for quantitatively assessing human text legibility in images and the methodology mainly involves *a posteriori* testing. In Section 5.2, on the other hand, questions about general (phenomenological) processes are asked when analyzing the responses obtained in a user study for coherence and random effects.

## 1.4 Structure of the Thesis

In Chapter 2, related work in the areas of multispectral imaging, photometric stereo and image quality assessment is discussed, whereby the focus lies on aspects relevant for the subsequent chapters.

Chapter 3 describes devices and procedures for multispectral imaging and photometric stereo, which are used for data acquisition throughout this work.

Chapter 4 continues with investigating aspects of photometric stereo that are relevant for the documentation and visualization of graphical heritage and demonstrates their application in a case study.

In Chapter 5, the problem of evaluating the quality of images of graphical heritage is addressed: after preliminary discussions, a novel dataset and evaluation framework for image quality estimators are introduced.

Chapter 6 concludes this work with a summary and critical discussion of the results presented.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

CHAPTER 2

# Related Work

Open your eyes, open your mind.

*Sandra Nasić*

This chapter reviews work in the areas of multispectral imaging, photometric stereo and image quality assessment, which is relevant for the topics treated in the subsequent chapters. For multispectral imaging, the focus lies on image acquisition systems for cultural heritage applications and the problem of chromatic aberrations. After describing the fundamentals of photometric stereo, generalizations to non-parallel lighting and non-Lambertian reflectance are reviewed, before turning to light source calibration and uncalibrated photometric stereo, as well as works on error analysis. Finally, approaches for image quality assessment are discussed, with a focus on images of writings.

## 2.1 Multispectral Imaging

The human eye is sensitive to electromagnetic radiation in a limited range; depending on the amount of radiation entering the retina and individual sensitivity of the observer, a lower bound between 360 nm and 400 nm and an upper bound between 760 nm and 830 nm are assumed in literature [Sli16, ST97]. Perceived colors result from radiation in this range that is emitted or reflected from objects and stimulates three different kinds of light sensitive cells in the observer's retina, each responding to a different range of wavelengths [ST97]. This trichromatic system gives rise to metamerism: two different spectra entering the retina invoke the same sense of color, if three cone types are stimulated with the same intensities [ST97]. The same is true for conventional RGB photographs, which record three color channels analogous to human vision [ST97].

Multispectral (MS) imaging extends the capabilities of human vision and conventional color photography by providing a finer sampling of the spectrum (*i.e.*, differentiating more

TU Bibliothek, WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

spectral bands) and/or including non-visible wavelengths such as ultraviolet (UV) and infrared (IR) [JDGT20]. Thus, it aids the accessing of graphical heritage that manifests in variations of reflectance spectra hard or impossible to distinguish with the naked eye or conventional digitization, for example: faded writings on ostraca [FSS+12] or parchment [HDF+15], undertexts of palimpsests [ECK11, Kno18], poorly preserved wall paintings [ACC+19, KDM+18] or underdrawings [FK06, GCL+19].

After briefly reviewing the different types of acquisition systems used in the context of graphical heritage, the remainder of this section focuses on specific challenges in MS imaging that are treated in this thesis.

### 2.1.1 Multispectral image acquisition

MS imaging aims at observing the interaction of surfaces with multiple specified wavebands. The approaches considered in the following rely on acquiring multiple images of the same scene, where for each image, the emission spectrum of the light source and/or the spectral response of the imaging device are adjusted - henceforth, these different images are referred to as *layers* of an MS image. *Hyperspectral* imaging devices that differentiate hundreds of wavebands by dispersing a single scan line across a 2D sensor and record a 2D surface via scanning motion [PCCS20] (and generally trade spectral resolution for spatial resolution [JDGT20]) are not considered here.

While it is commonly assumed that MS imaging has its origins in remote sensing [JDGT20, HS17, FK06, WCCM00], the first MS imaging systems for graphical heritage are documented at the beginning of the 20th century (although not labeled as such). In 1901, Pringsheim and Gradenwitz describe the enhancement of palimpsest undertexts by imaging a manuscript page with two different photographic plates: one sensitive to lower wavelengths (UV and blue), where both text layers are equally absorbent; and one sensitive to higher wavelengths, where the undertext is less absorbent than the overtext [PG01]. When overlaying the positive of the former and the negative of the latter image, the undertext stands out. The authors already mention the difficulties of producing perfectly congruent images, when elements in the optical path are altered [PG01]. In 1914, Kögel describes the combination of different photographic plates and optical filters for different use cases in palimpsest imaging [Kög14]. Among other things, he describes methods for UV reflectography and fluorescence imaging: with an electric arc lamp as a UV-rich light source, UV reflectography is realized by placing an optical filter in front of the camera (the author recommends a silver-coated glass element or a cuvette filled with a certain chemical) and using a UV-permissive quartz lens; hereby, the difficulties in focusing the image are mentioned. UV fluorescence images are achieved by dispersing the light source with a prism, such that only the UV part of the spectrum reaches the palimpsest [Kög14]; the author also describes the usefulness of this method for palimpsest photography, as the parchment background fluoresces strongly, while the erased writings do not.

After having established that neither MS imaging nor its application to graphical heritage are particularly novel, we now turn to more recent systems based on digital image sensors.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Among these, two fundamental approaches for the separation of wavebands are found in literature:

1. The object is illuminated with a broad-band light source and a waveband is selected from the reflected light by means of optical filters.

2. The object is illuminated with different narrow-band light sources, and the reflected light is recorded as it is.

In 1998, Baronti *et al.* describe a comprehensive system of the first category for analysis of paintings [BCLP98]. Broad-band illumination covering the visible and near-IR range is provided by two projectors with quartz tungsten halogen lamps. An eight-slot motorized filter wheel that can be equipped with various band-pass filters ranging between 420 nm and 1550 nm is placed in front of a PbO–PbS vidicon camera operating between 400 nm and 2200 nm [BCLP98]. The system also features a contrast-based autofocus system to handle focus loss in the near-IR range; for coping with the resulting misalignments between images, a 4-point image registration approach is used. The authors further describe the photometric calibration of the images using *Spectralon* reflectance standards and the use of principal component analysis for dimensionality reduction (*i.e.*, to condense the MS stack to few expressive images) [BCLP98].

Kleber *et al.* use a similar system for on-site imaging of degraded parchment manuscripts [KDL+08, GMLS11], but with a CCD camera (Hamamatsu C9300-124) ranging between ca. 330 nm and 1000 nm. They employ two halogen lamps and five band-pass filters in the visible and near-IR range; furthermore, UV tube lamps, in combination with a 400 nm short-pass and long-pass filter, are used to produce UV reflectography and fluorescence images, respectively. Additionally to the aforementioned configurations imaged with an achromatic camera, the authors describe the use of a conventional RGB DSLR camera for recording visible color images and color-resolved UV fluorescence images (depending on the active light source). Ware *et al.* employ yet a similar system (incandescent light sources and filters in front of a CCD camera) for the imaging of carbonized papyri, but describe the use of a tunable LCD filter as an alternative to a filter wheel [WCCM00].

Easton *et al.* use a system falling in the second category, where wavebands are separated by means of illumination [ECK11]. Thirteen spectral bands between 365 nm and 1050 nm are produced by different LEDs arranged in illumination panels; a 3-position filter wheel mounted in front of the CCD camera allows the recording of fluorescence images. Kleynhans *et al.* describe a low-cost variant of a MS document imaging system based on LED panels and a small achromatic camera, which is designed to be easily assembled, transported and used by scholars with limited technical background [KCM21]. Due to the lack of a filter wheel and lens characteristics, UV reflectography is not supported. At the time of writing, open source publication of component lists, building instruction and capturing software is in preparation.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

At the time of writing, commercial systems are available as well. *XpeCam* [PMd23] is a mobile and compact system for conservational fieldwork. The illumination devices house three lamps covering the UV, visible and IR ranges, and a 30-position filter wheel integrated with the 6.4 megapixel camera selects wavebands between 350 nm and 1200 nm; fluorescence imaging is not supported. On the other hand, the *PhaseOne Rainbow* system [CMDPD23] is a high-end solution designed for stationary use, with a 100 megapixel medium format camera and support for MS LED panels (similar to Easton *et al.* [EKC+10]) as well as broad-band light sources, optical filters and a variety of reflectance/luminescence imaging modes [DVC13].

### 2.1.2 Handling chromatic aberrations

The previous section has already touched on an inherent problem of MS imaging that is caused by optical dispersion: the refractive indices of materials depend on the wavelength of refracted light [ŠKB+10, MN13]. In photography, this effect manifests as chromatic aberrations, leading to both offsets between color channels in the image plane (lateral chromatic aberration) and offsets of their focal planes (longitudinal chromatic aberration) [JF06]. In the extreme case of MS images, where the range of utilized wavelengths typically ranges from UV to IR, the effect is pronounced. While it is possible to design lenses that minimize chromatic aberrations across the range of imaged wavelengths [HM63, MN13], it is not possible to completely eliminate aberrations for the whole optical spectrum utilized for MS imaging.

Lateral aberrations, whether caused by dispersion, application of optical filters [BSA08] or other influences, can be corrected in post-processing via conventional image registration procedures or by using prior information from dedicated geometric calibration [ŠKB+10]. On the other hand, information is lost irreversibly when channels are out of focus. Extensive research has been done on deblurring images by deconvolution [CE07, PF16]. While Heide *et al.* show that chromatic aberrations in images taken with simplistic lenses can be greatly reduced using such methods [HRH+13], frequencies that have been entirely eliminated by convolution (*i.e.*, multiplied by zero in frequency space) cannot be restored by deconvolution. It is therefore desirable to avoid focus shifts at acquisition time.

The problem of focus loss at non-visible wavelengths is reported in MS imaging literature and handled in different ways. Several systems proposed for imaging of historic documents rely on the use of apochromatic lenses corrected for the desired range of wavelengths. For example, Easton *et al.* and Mathys *et al.* report the use of a *CoastalOpt* 60mm UV-VIS-IR lens by *Jenoptik*, which is color corrected from 310 nm to 1100 nm [MJH19, EKC+10, EKCB18]. This lens is designed for use with full format (35 mm) sensors; when used with larger sensors, degraded quality must be expected close to the edges [EKCB18]. The authors use illumination wavelengths between 365 nm and 1050 nm and do not report focus problems. Kasotakis *et al.* describe a "custom designed" lens corrected for UV, IR and visible wavelengths, which they use for their imaging system in combination with a medium format sensor (49.1 mm × 36.8 mm) [KMB22]. Kleynhans *et al.* recommend a *Schneider Kreuznach Citrine* VIS-IR C-mount lens, color corrected from 400 nm to

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

1000 nm, for their low-cost system [KCM21]; however, imperfect focus is reported for wavelengths within this range.

When using a lens that introduces intolerable aberrations for wavelengths of interest, other correction strategies at imaging time are possible. For applications where automation of the imaging process is not envisaged, manually adjusting the focus for different wavelengths is reasonable [Kög14, DVC13]. Another straightforward idea is to autofocus the image prior to each shot. This approach is mentioned by Baronti *et al.* for images in the range of 420 nm to 1550 nm, but aside from mentioning a micromotor and an algorithm looking "for the maximum image contrast", implementation details are sparse [BCLP98]. Similarly, Papadakis *et al.* mention an unspecified "smart auto-focus mechanism", which, however is available for automated acquisition mode only in the range between 400 nm and 1000 nm, and is a limiting factor of acquisition duration [PMd23]. Rosenberger and Linß determine the focus shifts occurring in a filter-based MS microscopy system by measuring the linear movement necessary to achieve maximum focus; the focus shift is then compensated by adding correction plates of appropriate thicknesses to the filters [RL15].

Even if the longitudinal aberrations are corrected by any of the means described above, lateral aberrations can still be present or even amplified, when adjustments to the focus are made during imaging [BCLP98, RL15]. The resulting misalignments between spectral layers must be corrected in post-processing, either by explicitly modeling the displacements occurring [BSA08, RL15], or via image registration [LDSM08, MGC$^+$13, CDL15].

## 2.2 Photometric Stereo

Photometric Stereo (PS) is a method for reconstructing surface orientation and depth from a set of images with constant camera parameters and varying lighting directions [Woo80]. In comparison to other 3D acquisition methods (*e.g.*, structured light scanning or photogrammetry), PS is especially efficient for the acquisition of local/high-frequency surface details while being prone to global/low-frequency distortions [HW11]; this manifests in typical application scenarios like inspecting the waviness of planed wood surfaces [JYP07], checking carbon fiber orientations [TPSE13] or separating handwriting from printed background text based on the impressions left by the pen [MC03]. Due to these merits paired with cheap and flexible acquisition procedures [VBMR15], it is also used in documenting depth-variation-based graphical heritage, such as historical coins [HMV18, BZS18], engravings and reliefs in rock [DMR$^+$15, VBMR15] or surface structures in medieval manuscripts [VHW$^+$18]. In the following, the fundamentals of PS are discussed, followed by extensions and aspects of the method that are relevant for the work presented in this thesis.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

### 2.2.1 Original formulation

In his seminal paper [Woo80], Woodham formulates the first version of PS based on the Lambertian illumination model

$$I = \rho \ cos\theta, \tag{2.1}$$

where $I$ is the portion of reflected light, depending on surface reflectance/albedo $\rho$ and the angle between surface normal vector and incident light direction $\theta$, $0 \leq \theta \leq \frac{\pi}{2}$; an intuitive illustration is given in Figure 2.1a. Expressed via surface normal vector $\vec{n} = [n_x, n_y, n_z]'$ and light direction $\vec{s} = [s_x, s_y, s_z]$, $|\vec{n}| = |\vec{s}| = 1$, this results in

$$I = \rho \ \vec{s} \cdot \vec{n}. \tag{2.2}$$

If the reflected radiations $I_1, I_2, I_3$ resulting from three known light directions $\vec{s}_1, \vec{s}_2, \vec{s}_3$ are measured, the linear system

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \rho \begin{bmatrix} \vec{s}_1 \\ \vec{s}_2 \\ \vec{s}_3 \end{bmatrix} \vec{n} = \rho \ \mathbf{S} \ \vec{n} \tag{2.3}$$

allows the determination of albedo-scaled surface normals via:

$$\rho \ \vec{n} = \mathbf{S}^{-1} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}; \tag{2.4}$$

see Figure 2.1b for illustration. The linear system is only solvable if the light direction matrix $\mathbf{S}$ has an inverse and thus the light directions are linearly independent. For more than three input images, the system is over-determined and solvable *e.g.* with least-squares optimization [WGS+11].

Aside from a Lambertian reflectance of imaged surfaces, this straightforward formulation of PS makes further simplifying assumptions, like parallel illumination and the absence of cast shadows. To account for deviations from this idealized model in real-world acquisition scenarios, numerous extensions have been proposed [HW11]. In the following, only the aspects most relevant for this thesis are reviewed.

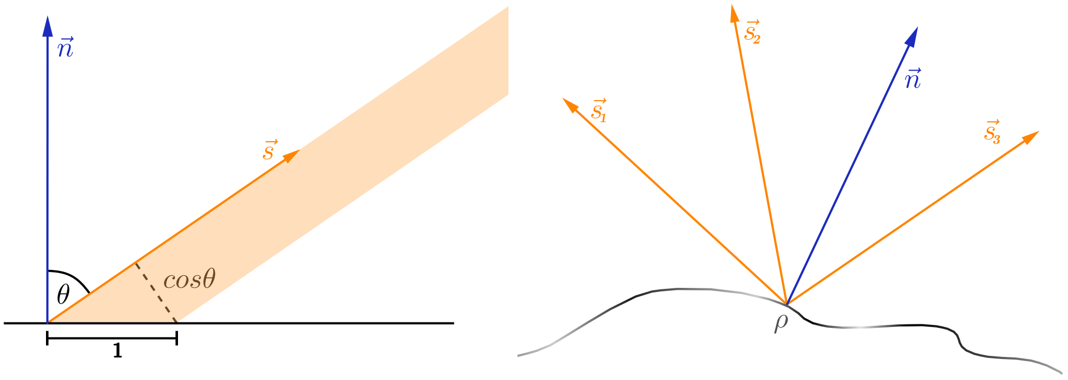
### 2.2.2 Non-parallel lighting

In the original formulation [Woo80], parallel lighting (*e.g.*, by infinitely distant light sources) is assumed; thus, the light direction and intensity is considered equal for each surface point imaged, regardless of its location in space. This enables a straightforward estimation of surface orientations [Woo80] and the computation of depth in an independent step via integration [SCS90].

Iwahori *et al.* were among the first to adapt PS to point light sources, proposing the simultaneous estimation of depth and surface orientation with a set of non-linear

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b)

Figure 2.1: (a) The amount of light reaching a surface patch is proportional to the cosine of the incident angle $\theta$; in the Lambertian illumination model, this light is reflected in all directions equally, scaled by a reflectance factor (albedo). (b) The basic idea of PS: normal vector $\vec{n}$ and albedo $\rho$ are reconstructed from image intensities measured under illumination directions $\vec{s}_1$, $\vec{s}_2$, $\vec{s}_3$.

equations [ISI90]. Clark reformulates the problem by considering movements between light positions and the resulting relative changes in intensity, thereby arriving at linear equations [Cla92]. Both Clark and Iwahori *et al.* assume isotropic quadratic light attenuation. Later works are based on a more complete lighting model, including both attenuation due to distance and radial attenuation (depending on the deviation from the principal direction of a light source); Mecca *et al.* and Quéau *et al.* propose solutions via solving partial differential equations [MWBK14, QDW+18]. Another class of solutions is based on the alternating estimation of normals and depth [NS16, XNSW19]. The latest approaches use deep neural networks [LBMC20, LSJ22]. A state-of-the-art method by Lichy *et al.* [LSJ22], that stands out for being faster and more memory-efficient than its competitors, still needs 4 GB of CPU memory and 12 GB of GPU memory for processing 1024×786 input images.

Despite the proposed solutions, authors continue to use parallel lighting PS although their experimental setups make use of point light sources. This is done to avoid unnecessary complexity in works treating other extensions of PS like unknown light directions [PF13, QLD15] or non-Lambertian surfaces [SSS+07, WGS+11], but also in practical applications, *e.g.*, discrimination of handwritten and printed text based on impressions in the paper [MC03], detection of production flaws in ceramic tiles [FSSM05], measuring waviness of planed wood surfaces [JYP07], fingerprint scanning [XSC13], or capturing the reliefs on historical coins [BZS18].

In literature, distinctions are made between distant-light scenarios (where the parallel light assumption is feasible) and near-light scenarios (where it is not), depending on the size of the imaged scene/object in relation to the distance of light sources - view Figure 2.2 for an intuitive understanding. However, this distinction is vague. Authors

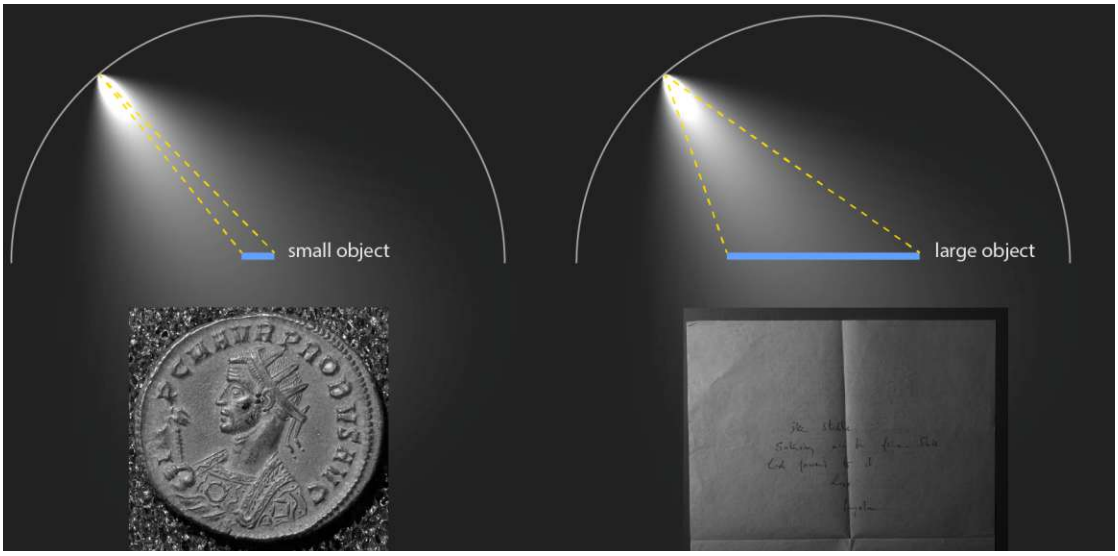


Figure 2.2: The larger the imaged object relative to the distance of light sources, the larger are variations of incident light direction and intensity across its surface. The example images show a roman coin (Ø 22 mm) and half of an A4 page (148 mm × 210 mm), both illuminated by an LED placed ca. 450 mm from the object center.

state that in a near-light scenario, lights are "near the object" [Cla92], that the "distant light assumption is a reasonable approximation as long as the dimensions of the scene are much smaller than the distance of the light sources" [PF14], that "the distant light source assumption fails when the object-to-light distance becomes small" [LND18], *etc.* Others are more precise, claiming that the parallel light assumption is problematic "when the object size is comparable in magnitude to the light separation" [ASSS14] or that "illumination is nearly parallel only when the distance between a lighting source and the object is more than 10 times of the object's dimensions" [XCW15]. This last quite specific claim, however, comes without a justification other than a citation of Woodham's seminal paper [Woo80], which indeed does not comment on this topic.

In order to circumvent the complications introduced by point lighting models, several heuristics are proposed to improve reconstruction quality when using parallel lighting models. Fan *et al.* use their error estimations to remove low-frequency reconstruction errors in the depth maps in a post-processing step [FQW$^+$17]. Sun *et al.* propose to apply a conventional flatfield correction on the input images to compensate for uneven lighting [SSSF13]. This heuristic is only applicable if calibration surface and reconstructed surface are similar. Another approach consists of the combination of PS with low-resolution reference depth data, such as sparse control points [HK04] or structured light scans [LTBB10]; thus, accurate low-frequency shape is augmented with high-frequency details from PS. Guo *et al.* base their uncalibrated PS method on the assumption that even in a general illumination setting, illumination is approximately parallel within small surface patches [GMS$^+$21]. They solve uncalibrated PS locally for

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

each patch and handle the inherent ambiguities with a Markov Random Field optimization. The method is demonstrated on input images of 160 × 160 pixels and surface patches of 3 × 3 pixels; non-uniform albedo within a surface patch negatively impacts reconstruction quality [GMS+21].

### 2.2.3 Non-Lambertian reflectance and shadows

Further simplifications of original PS are the assumption of purely diffuse (Lambertian) surfaces and the absence of shadows. Some works tackle this problem with more complex reflectance models [NIK90, AWL13]. Another approach is to remove outliers of the assumed model (*e.g.*, the Lambertian one) from the input data as a pre-processing step. Wu *et al.* provide a method to remove shadows and specularities from the input data by low-rank matrix recovery [WGS+11].

### 2.2.4 Light source calibration

Traditional PS approaches require knowledge of the direction of incident light for each input image [Woo80]. In the following, methods for determining light source directions/positions from calibration images are discussed. A widely used general approach proposed by Powell *et al.* involves placing two or more reflective spheres in the scene and triangulating the light source positions based on specular highlights visible in the respective images [PSG01]. Masselus *et al.* demonstrate a similar approach based on diffuse spheres: instead of specular highlights, intensities across the sphere surface are utilized to estimate lighting directions with the Lambertian reflectance model [MDA02]. Ackermann *et al.* extend the reflective spheres method by optimizing the back-projection errors of estimated light source positions, similarly to the optimization of camera parameters in structure from motion methods [AG15]; they also respect that under perspective projection, spheres are generally mapped to ellipses. Nie *et al.* also use reflective spheres for determining light source positions and additionally estimate the principal axis of non-isotropic light sources from brightness value distributions on a planar surface [NSJZ16]. Note that when lighting is parallel, the estimation of a light direction instead of a position is required; for this, only one sphere is enough [DY13]. An interesting alternative method is proposed by Santo *et al.* they estimate light source positions from the movements of shadows of pin needles stuck in a planar surface [SWL+20].

For methods utilizing reflective spheres, the segmenting of these spheres in the input images is a surprisingly non-trivial sub-problem, mainly because cast shadows and reflections lead to unreliable edges [LBT+17]. Powell *et al.* bypass the problem by using spheres that are painted in a bright diffuse color on one hemisphere and mounted on sticks, and take different images for sphere and highlight detection, with the spheres rotated in between [PSG01]. Ackermann *et al.* fit ellipses manually and just note that the “procedure could be automated by first segmenting the sphere,...” [AFG13]. Nie *et al.* claim to segment the spheres “via Canny operator” without further explanation [NSJZ16].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Only Liao *et al.* elaborate on a heuristic for sphere segmentation that is based on the median gradients computed for multiple input images for edge detection [LBT+17].

### 2.2.5 Uncalibrated Photometric Stereo

Uncalibrated PS refers to methods that estimate surface normals without knowledge of the light directions corresponding to the input images. Belhumeur *et al.* show that under the assumption of an orthogonal projection (which is made across the majority of PS literature) and without further restrictions, this problem is ill-posed and can only be solved up to generalized bas-relief transformations $\overline{f}(x, y) = \lambda f(x, y) + \mu x + \nu y$, where $f(x, y)$ denotes the height at position $(x, y)$ [BKY99]. Approaches to estimate the parameters $\lambda$, $\mu$ and $\nu$ of the generalized bas-relief transformation rely on regularity measures. Alldrin *et al.* propose the minimization of the entropy of the albedo map [AMK07]. Queau *et al.*, similarly, employ the total variation of the resulting normal- or depth field as a regularity measure, thereby taking into account spatial relations between surface points [QLD15].

### 2.2.6 Error analysis

Several works investigate the influence of uncertainties/errors in the lighting model on reconstruction quality. Among the first are Ray *et al.*, who identify potential sources of errors and derive the sensitivity of reconstructed surface normals with respect to errors in measured intensities and calibrated light directions [RBK83]. Their analysis is limited to a setup with only three light sources, and validated by showing that reconstruction errors on a diffuse sphere are explained by their error predictions up to 5°. The influence of non-parallel lighting is not analyzed directly, but is said to be reducible to spatially varying direction errors. Jiang and Bunke reproduce the theoretical sensitivity analysis of Ray *et al.* with a more compact formulation [JB91]. They summarize the results of the analysis by stating that a 1° error in reconstructed surface orientation is introduced by a 1% error in a measured intensity or a 1° error in a light direction. Schlüns finds that for three-light PS, the possible normal errors introduced by intensity measurement errors of given magnitude are described as an ellipsoid [Sch97]. Drbohlav and Chantler investigate optimal light configurations with respect to robustness against image intensity noise. They find that three light sources are optimally arranged if orthogonal to each other, and that the optimal slant angle is at $\arctan \sqrt{2} \approx 54.74°$ [DC05]. Spence and Chantler experimentally arrive at similar results [SC06]. Sun *et al.* examine the uncertainty in reconstructed normals introduced by uncertainties in light intensities; they conclude that for a three-light setup, a 1% variance in intensity introduces errors of 0.5° to 3.5° in the surface normal direction, depending on the relative arrangement of light sources [SSSF07]. Kobayashi *et al.* find that light direction calibration errors, as well as resulting normal reconstruction errors, are described by a Fisher distribution [KOMS11]. Fan *et al.* find that assuming parallel lighting in a point light setup introduces linear errors in gradients and quadratic errors in the depth map, as a function of image position [FQW+17]. They validate the error model only on quasi-flat surfaces and using a circular arrangement of light sources. Chen *et al.* analyze the impact of light source calibration errors on normal

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

reconstruction errors in point light setups [CRZ+22], finding that deviations in light source positions and non-uniformity of illuminants (*i.e.*, spatially varying intensities) cause the greatest errors. Like previous authors, they only consider setups with three light sources.

### 2.2.7 The advent of deep learning

The works discussed above largely address the PS problem analytically, based on physics-inspired models. Like in other areas of computer vision, this general approach is increasingly displaced by learning-based methods. Indeed, recent works excel in challenging conditions like unknown and general lighting as well as complex surface reflectance and the presence of shadows [CHS+19, GMS+21, LSJ22, Ike23b, LZS+23]. While producing impressive results, the hardware requirements of these state-of-the-art methods can be prohibitive for practical applications. For example, the method by Ikehata [Ike23b] requires 40 GB of GPU memory solving PS with ten input images of 2048 × 2048 pixels [Ike23a] (although with an alternative implementation, the consumption can be reduced to “only” 10 GB).

## 2.3 Quality assessment

If the aim of an imaging or image processing method is to generate images that are useful for the discovery and analysis of graphical heritage by human researchers (cf. Section 1.1.2), assessing the quality of such images (and thus the quality of the generating method) is not straightforward. Consequently, the evaluation of proposed approaches is commonly based on expert ratings, the demonstration on selected examples, or case studies [ECK11, HDF+15, Min18, PDG+17, STB07]. This practice is unfavorable for the research field: it does not allow for an automated evaluation on large public datasets, such that an objective comparison of different approaches is impeded. In the following, approaches for quantitatively evaluating images of graphical heritage are reviewed. The focus lies on *legibility*, *i.e.*, the aptness of an image to convey textual contents - however, many of the works mentioned are generalizable to pictorial contents as well.

In general Image Quality Assessment (IQA), methods are typically classified with respect to the additional information necessary to make a quality judgment [MMB12, MS15, BMM+18]: Full/reduced reference approaches require a reference image of optimal quality or a reduced representation, such as a ground truth segmentation, for operation. A typical use case for full-reference methods is the evaluation of lossy image compression, where the uncompressed version serves as a reference [APY16, WBS+04]. No-reference or blind methods, on the other hand, operate on the input image only. Such methods might be trained and tested using reference information but do not require this information in production mode. This means that they are applicable to arbitrary images and usable in a wide range of scenarios where no reference is available. Regarding the assessment of legibility as a variation of general IQA with a different definition of “quality”, these

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

categories are also used for a coarse structuring of corresponding approaches discussed below.

### 2.3.1 Full/reduced reference approaches

Giacometti *et al.* created an MS image dataset of manuscript patches before and after artificial degradation [GCM+17]. This allows the quantitative assessment of digital restoration methods by comparing the results to the non-degraded originals. The authors thus follow a full reference approach [MS15] - the only one described in this section. As for historical manuscripts and other archaeological objects an image of its non-degraded state is usually not available, the application is limited to expensively created test datasets.

Arsene *et al.* have specifically evaluated the application of dimensionality reduction methods for text restoration based on MS manuscript images [ACD18]. Their evaluation is based on both expert ratings (given by seven scholars) and cluster separability metrics (Davies-Bouldin Index [DB79] and Dunn-Index [Dun73]). The clusters tested for separability are defined as manually selected pixels from foreground and background. In their experiments, the cluster separability metrics could correctly identify the image variants rated best by the human judges; apart from that, cluster scores and human assessments did not correlate well. The authors acknowledge this and claim that visual assessment by philologists is still the standard for evaluating readability enhancement methods.

A natural assumption is that the quality of an image with regard to readability is strongly connected to its contrast; this is problematic, however, as high contrast can be found in substrate variations (*e.g.*, stains) or noise. Furthermore, the nominal contrast of an image can be increased by simple intensity transformations. Shaus *et al.* introduce the metric of *potential contrast* and suggest its application as a quality metric for images of ostraca [SFGST17, FGSS+12]. The metric measures the contrast achievable between foreground and background under arbitrary grayscale transformations. Similar to the previous method, it relies on user-defined foreground and background pixels. The correspondence to human assessments is not evaluated.

For restoration approaches producing binary images, evaluation is typically done by quantitative comparison to a ground truth segmentation [SON19]. This comparison is straightforward with standard error metrics such as F-score or peak signal-to-noise ratio [PZK+19]. Datasets with ground truth segmentations of historical manuscripts have been published, with both RGB images [GNP09, PZK+19] and MS images [HNM+15, HBS19] as inputs.

A quantitative metric especially for evaluating document image enhancement methods is the performance of Optical Character Recognition (OCR) on the enhanced images [LSDS11, HDS14]. This approach addresses the property of *legibility* more directly than the approaches mentioned before and is a reasonable choice when the purpose of an image lies in the subsequent automated transcription, rather than in the delivery to a human observer. The correspondence between machine legibility and human legibility, however, is not straightforward [OAA12]; a fact that is exploited, for example, by

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

*CAPTCHA* authentication methods [XLL20]. A further problem with this method is the strong dependence of OCR systems on the scripts and languages they have been trained on [LLS20]; this is especially problematic when considering ancient documents using rare historical scripts. For example, Hollaus *et al.* evaluate their work on Glagolitic script using a custom OCR system that has been trained for Glagolitic script only [HDS14].

### 2.3.2 No-reference/blind approaches

General blind IQA approaches that are not limited to a certain type of distortion typically employ machine learning in some form [MS15, BCNS18]. While early methods based on natural scene statistics, such as DIIVINE [MB11] or BRISQUE [MMB12], are largely hand-crafted and just "calibrated" on a training dataset, later publications make heavy use of convolutional neural networks [BCNS18, BMM$^+$18, LW18, KYLD14, LVDWB17]. No-reference IQA has been used to select optimal parameters for de-noising [MMB12, ZMM10] and artifact removal in image synthesis [AKMS12]. General IQA metrics such as BRISQUE [MMB12] are used as reference implementations for specialized document IQA methods, showing comparable performance [GC16, SNAMC18]. Ye *et al.* propose a general IQA method based on filter learning and specifically demonstrate it on the assessment of document image quality [YKKD13]. Garg and Chaudhury later use the same method for automated parameter tuning for document image enhancement [GC16]. Both publications validate their document quality estimates via OCR performance on machine-written documents. The correlation of the predictions with human legibility or perceived quality is not addressed.

A similar situation is found in dedicated *document IQA* literature [YD13a]: Quality is mostly defined in terms of OCR accuracy and the proposed methods are trained and tested accordingly [YD13a, LZQ18, LZQ19, LD19]. Document IQA thus amounts to OCR performance prediction. In Section 2.3.1, it was argued that OCR performance cannot be directly used as a human legibility estimator; similarly, OCR performance *predictors* cannot be directly used as human legibility estimators [OAA12].

Few approaches explicitly target the estimation of human legibility or perceptual quality in documents. Stommel and Frieder propose an automatic legibility estimation for binarized historical documents, intended for unsupervised parameter tuning of binarization algorithms [SF11]. The method relies on OCR features computed for small image patches. Based on these features and subjective ratings (good, medium or bad legibility) an SVM classifier is trained. Applied to a whole document in a sliding window manner, the method produces a spatial map of legibility.

Obafemi and Agam propose a document IQA estimator based on classical feature extraction and a neural network classifier. The method operates on binary images of segmented characters. They show that their system can be trained to estimate both human ratings and OCR accuracy; however, it must be trained for each of these use cases separately. The approach is demonstrated on machine-written text only.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Shahkolaei *et al.* created a dataset for quality assessment of subjectively rated ancient document images [SNAMC18]. The dataset is based on 177 RGB images of Arabic manuscript pages. Mean opinion scores for the quality of the pages were obtained in a study where 28 students of technical subjects rated the images in a pair comparison mode. The images were compared with respect to not further specified “quality” on a full-page basis. This dataset is used to train and test a novel method for objective quality assessment of degraded manuscript images, which uses support vector regression on Gabor filter responses [SNAMC18].

Lastly, I would like to address a class of methods that, to the best of my knowledge, was not used for legibility estimation before, but is worth investigating. While the final transcription accuracy of OCR systems was argued to be problematic for the purpose, early stages of OCR pipelines could be useful for blind legibility estimation: before text can be transcribed, it must be localized in the input image. In the domain of document image processing, text line detection and layout analysis are typical processing stages [BKJ+17, GLD+18, GLS+19]. When working with natural images (*i.e.*, trying to detect text “in the wild”), the problem is referred to as scene text detection [NM12, LSB+17, ZYW+17]. For both cases, systems that are largely independent of script and language are described [DKS13, GLS+19, RKC14, DSR+17].

## 2.4 Summary

Compared to human vision and conventional photography, MS imaging increases spectral range and resolution; early versions of the technology were used at the turn of the 20th century for the restoration of palimpsests. Modern systems achieve the separation of wavebands by means of optical filters and/or narrow-band light sources. Due to the extended spectral range, chromatic aberrations are pronounced in MS imaging. While lateral aberrations can be compensated in post-processing, longitudinal aberrations leading to focus loss must be avoided while imaging. Solutions for avoiding focus loss include the use of special lenses, auto-focusing before each shot and the introduction of waveband-specific corrective optical elements.

PS is a method for reconstructing surface orientation and albedo from a set of images captured under different lighting directions. In the original formulation, the Lambertian reflectance model and parallel lighting are assumed; the latter assumption is violated in point light setups. The smaller the distance of light sources to the imaged object compared to object size, the larger the resulting errors - however, the exact nature of this dependence is not researched. Various approaches are proposed for generalizing PS to different reflectance models, non-parallel or unknown lighting, with the latest methods being based on machine learning. Compared to the original formulation, however, these extensions are generally more expensive with respect to runtime and memory consumption.

Objectively evaluating the quality of graphical heritage imagery for human viewers is an open problem. When viewing the problem as a special case of general IQA, solutions

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

are classified in full-reference, reduced-reference or no-reference methods, depending on the additional information required to make a judgment. In the full-reference class, the only known approach involves datasets of artificially degraded objects. Reduced-reference approaches include comparing segmented/binarized images to a ground truth or measuring OCR performance on images of writing. Research on no-reference approaches, which would be preferable due to their general applicability outside of dedicated datasets, is sparse.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

CHAPTER 3

# Image acquisition and calibration

Like a rainbow in the dark, yeah.

*Ronnie James Dio*

This chapter describes devices and procedures for Multispectral (MS) imaging and Photometric Stereo (PS), which are used for data acquisition throughout this work. It starts with a hardware system originally designed for the imaging of historical manuscripts and continues with an approach for compensating focus shifts in MS imaging. Finally, practice-proven tools and heuristics for fine-registration of MS images and light position calibration for PS are described.

## 3.1 A mobile image acquisition system

The mobile MS imaging system described in this section is based on a system developed by Kleber *et al.* [KDL$^+$08, GMLS11] and later upgraded by Hollaus *et al.* with MS LED panels for illumination [HS17]. The third iteration, which is developed and used in the context of this work, features upgrades on camera, filter wheel and automation software, as well as an illumination dome for the acquisition of PS images.

The updated camera system consists of a *PhaseOne IQ260 Achromatic* camera back, a *Phase One XF* camera body and a *Schneider Kreuznach 120mm LS f/4.0 Macro* Lens. The change from the previously used Hamamatsu C9300-124, which indeed offers a better spectral response in the ultraviolet (UV) range, is mainly motivated by improvements with respect to spatial and photometric resolution: The medium format sensor of the *IQ260 Achromatic* (8964 x 6716 pixels on 53.7 x 40.3 mm) allows the adequate sampling of an object surface with fewer images, thereby accelerating the acquisition process; 16-bit raw data provide more information for subsequent processing and analyses.

| | *Hamamatsu C9300-124* | *Nikon D6* | *Phase One IQ260* |
|---|---|---|---|
| *image size* | 4000 x 2672 (ca. 10 MP) | 5568 x 3712 (ca. 20 MP) | 8964 x 6716 (ca. 60 MP) |
| *sensor size* | 36 x 24 mm | 35.9 x 23.9 mm | 53.7 x 40.3 mm |
| *photometric resolution* | 12 bit | 14 bit | 16 bit |
| *bayer filter* | no | yes | no |
| *IR blocking filter* | no | yes | no |
| *sensor cooling* | yes | no | no |
| *connectivity* | frame grabber card | USB-C | USB 3.0 |
| *max. exposure time* | 1 s | - | 2 m |
| *autofocus* | no | yes | yes |

Table 3.1: Comparison of the specifications of a scientific camera (Hamamatsu C9300-124, used by Kleber*et al.* [KDL$^+$08]), a state-of-the-art DSLR (Nikon D6) and the medium format camera used in this work (Phase One IQ260).

Secondary advantages lie in remote control via USB 3.0 and a manufacturer-provided API (without requiring a dedicated PC card), longer maximum exposure times (through which the inferior UV performance can be compensated) and support for auto focus. An achromatic camera is used because RGB color information is not required for the applications considered in this work, such that the presence of a Bayer filter would only necessitate additional interpolation operations on the raw data. Table 3.1 gives an exemplary comparison between a scientific camera, a state-of-the art DSLR camera and the achromatic medium format camera used in this work.

The *Schneider Kreuznach 120mm LS f/4.0 Macro* (roughly corresponding to a 75mm lens with a full-format sensor) is developed for Phase One cameras, such that the image quality matches the high-resolution sensors. A 1:1 magnification ratio is achieved at its minimum focal distance of 370mm; an area of an A4 page is fitted in an image at approximately 900mm imaging distance, with an magnification ratio of ca. 1:6. A geometric calibration of the camera system was performed in order to quantify the distortion characteristics of the lens, using the *MATLAB Camera Calibrator* [The21]. Following the definitions of Heikkilä and Silvén [HS97], radial and tangential distortion coefficients of $[k_1, k_2, p_1, p_2] = [0.0661, -0.5059, -0.0029, -0.0041]$ were determined; this leads to a maximum displacement of 30 pixels, occurring in the top left corner of the image. In the context of MS imaging, a major drawback of the lens is that it does not correct for chromatic aberrations in the UV and IR range; a method to overcome this problem is described in Section 3.2.

Illumination for MS imaging is provided by LED panels capable of producing 11 narrowband spectra from UV to near infrared (IR); the nominal and measured peak wavelengths of the built-in LED types are given in Table 3.2. In order to separate UV-induced fluorescence from reflected UV radiation, a motorized filter wheel (*Optec IFW3*) equipped with a UV blocking filter and a UV pass filter is mounted in front of the camera lens. Figure 3.1b shows the spectral response curve of the camera sensor, transmission spectra of optical filters, and the measured emission spectra of LED panels. Note the drop in sensor sensitivity in the UV range, which leads to long exposure times in UV reflectography. The

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

| *Nominal wavelength (nm)* | 365 | 450 | 465 | 505 | 535 | 570 | 625 | 700 | 780 | 870 | 940 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Peak wavelength (nm)* | 368 | 444 | 470 | 494 | 518 | 599 | 640 | 706 | 771 | 857 | 927 |
| *Description* | UV | royal blue | blue | cyan | green | amber | red | IR 1 | IR 2 | IR 3 | IR 4 |

Table 3.2: Overview of waveband emitted by the MS LED panels.

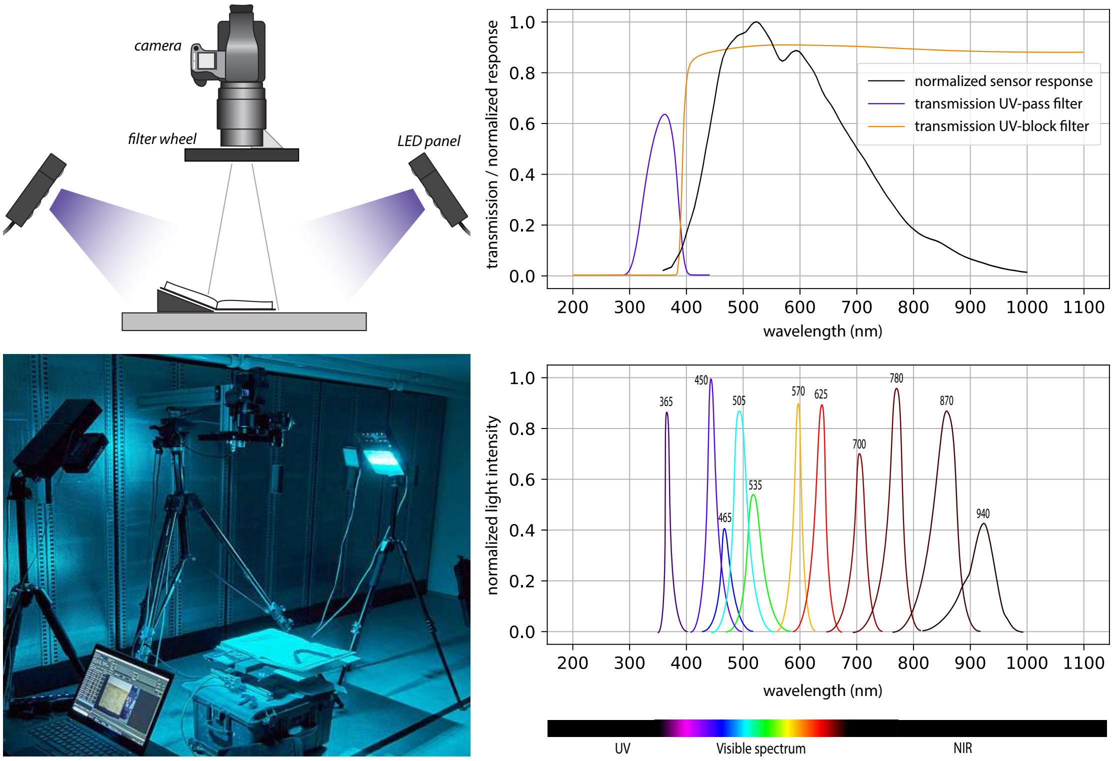
alignment of individual components is shown in Figure 3.1a [1]: the camera with filter wheel attached is positioned such that the principal axis is approximately perpendicular to the imaged surface, at a suitable distance (a *quasi-flat* surface, like a page of a manuscript, is assumed here). LED panels are placed symmetrically to the principal axis, such that the imaged surface is illuminated as evenly and shadow-free as possible. All components are connected to a (laptop) computer via USB and controlled by a custom acquisition software.

Acquisition of PS images is achieved by modifying the system described above with an alternative illumination device: 54 white LEDs (270 lm, half angle 65°) are mounted on a hemispherical dome with a radius of ca. 450 mm, while the camera is placed above an opening at the top (see Figure 3.2a). A skeleton made of spring steel rods was chosen over a closed dome for ease of object manipulation (*e.g.*, for turning pages of a codex) and transportability. The downsides of this choice are the restrictions to the arrangement of light sources and the need for a fully darkened work environment; however, if the environment cannot be darkened, usage in combination with a cloth cover is possible (see Figure 3.2b). The lights are arranged in 5 horizontal circles with twelve lights on the lower four circles and six lights on the uppermost (see Figure 3.2c). The LEDs are driven by a custom-built *Arduino*-based controller, which is operated by the acquisition software. As camera and light sources are not fixed relative to each other, light sources must be calibrated for each individual setup of the system - a convenient and straightforward approach is described in Section 3.3.

Figure 3.3 shows the graphical interface of the custom acquisition software[1]. It allows manual control of all hardware components, as well as the automatic execution of image series. For MS imaging, the user can define a custom list of image configurations, where for each shot, camera parameters, lighting and filters are specified. Such a list can be saved as an XML file for documentation and re-use. For PS, one image is taken for each LED, using manual camera parameters. During automatic acquisition (MS imaging or PS), lighting and filter configuration are encoded in the file name and camera parameters in the metadata of the raw file. Manual quality control is enabled by continuously displaying the latest recorded image. The software is written in C++ using the *Qt* framework; MS LEDs, filter wheel and PS illumination are controlled via commands over a serial interface, and the camera via a software development kit provided by the

[1] In Figure 3.1a, additional devices for the simultaneous acquisition of conventional color images are present: an RGB camera next to the main camera, white LED panels mounted on the MS panels, and a linear positioning unit to automatically move the object between the two cameras. Corresponding functionalities are also visible in Figure 3.3. These extensions, however, are not relevant for this work and thus not further described.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



(a) (b)

Figure 3.1: Multispectral image acquisition: (a) the hardware setup as a schematic and during an imaging campaign at the Klosterneuburg Abbey Library. (b) top: camera sensor response and transmission spectra of optical filters; bottom: emission spectra of the illumination system, whereby the peak annotations indicate the nominal wavelengths of the corresponding LEDs.

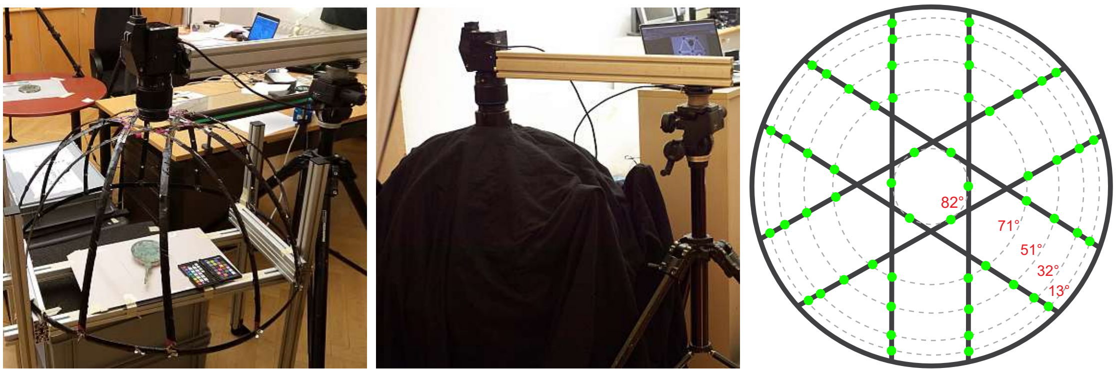
manufacturer. The focus shift correction for MS imaging described in the next section is integrated into the system as well.

At the time of writing, the acquisition system has been used for 22 on-site acquisition campaigns - from libraries, museums and archives in Austria and other European countries (Germany, Czech Republic, Bulgaria, Ukraine, Italy, Vatican) up to a burial cave in Egypt [KDM+18]. Multispectral images and/or PS reconstructions were acquired of more than 100 objects, and the results are used in several humanist or interdisciplinary publications [KDM+18, RBC+19, CMJ+22, CPB+22, FBF23, BFM23].

## 3.2 Focus shift correction in multispectral imaging

Chromatic aberrations, caused by wavelength-dependent refraction angles in optical elements, are an inherent challenge of MS imaging [ŠKB+10, MN13]. While lateral

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



(a) (b) (c)

Figure 3.2: PS acquisition setup. (a) arrangement of camera, illumination and imaged object, during an imaging campaign at the Kunsthistorisches Museum Wien; (b) with cover, when imaging environment cannot be darkened, here shown at Schloss Eggenberg, Joanneum Graz. (c) a schematic top view, where light source positions are shown in green and light sources located at the same elevation angle w.r.t the dome center are connected with a dashed line.

aberrations (displacements in image spaces) can be treated in post-processing, longitudinal aberrations (displacements of the focal plane) must be avoided during image acquisition, as they lead to out-of-focus images and thus irreversible data loss.

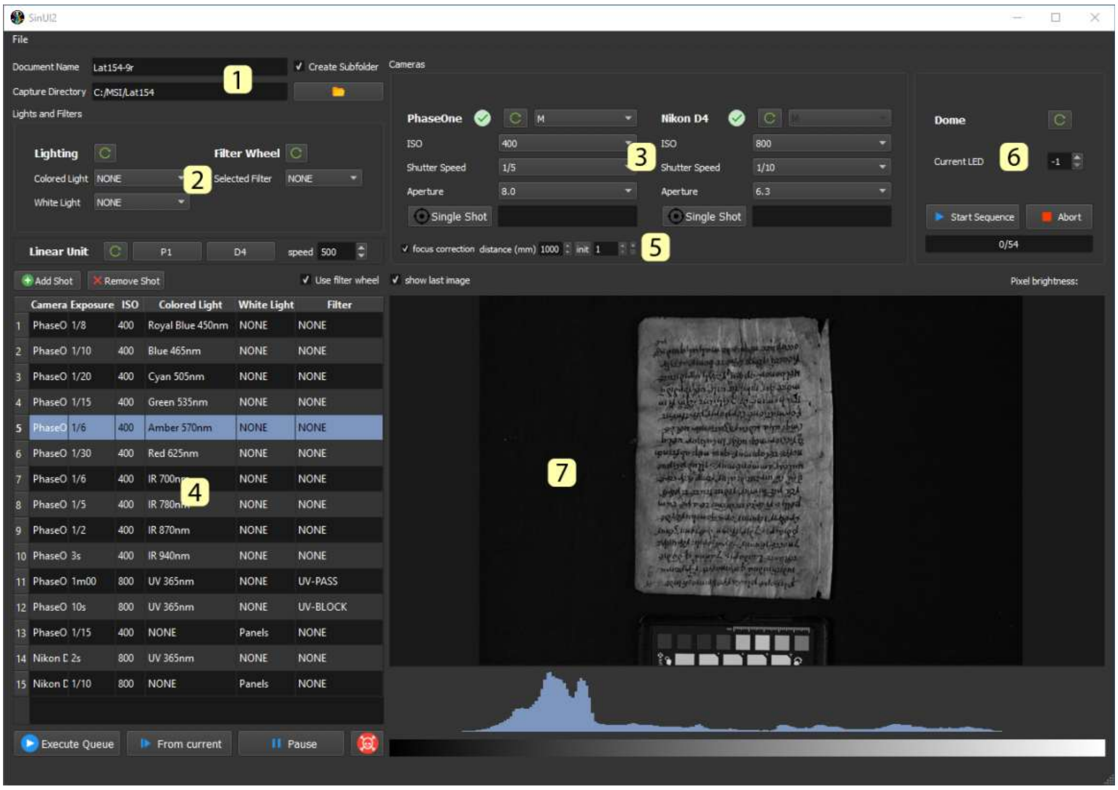
This section describes an approach for preventing such focus shifts at acquisition time, whenever the color correction capabilities of a lens are not sufficient for a given application. A model for the distance and wavelength dependent focus shift behavior of a lens is determined via a calibration procedure, and then used to inform mechanical adjustments during imaging. Integrated to the MS imaging system described in the previous section, the approach is shown to enable the acquisition of in-focus images in non-visible wavelengths using a lens designed for the visible spectrum only.

### 3.2.1 Calibration Procedure

Apart from the properties of the lens (which is here viewed as a black box), the magnitude of focus shift $\Delta$ with respect to a reference wavelength $\lambda_0$ depends on both the employed wavelength $\lambda$ and the distance $d$ between camera and object. Thus, the proposed calibration procedure aims at producing a function $\Delta(\lambda, d)$ describing the behavior of a given lens without any knowledge of its inner structure. We find this function by fitting to a set of empirical measurements.

To measure the focus shift for a given configuration $(\lambda, d)$, a calibration object is used. It consists of a target plane that is placed perpendicular to the principal axis of the camera, and a ruler plane, which is tilted by 45° with respect to the target plane (see top right of Figure 3.4 for illustration). Thus, when viewed from the camera, a known length $l$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



Document Name: Lat154-9r

Capture Directory: C:/MSI/Lat154

☑ Create Subfolder

Colored Light: NONE

White Light: NONE

Selected Filter: NONE

speed: 500

☑ Use filter wheel

☑ show last image

PhaseOne: M

ISO: 400

Shutter Speed: 1/5

Aperture: 8.0

Nikon D4:

ISO: 800

Shutter Speed: 1/10

Aperture: 6.3

☑ focus correction distance (mm) 1000 init 1

Current LED: -1

0/54

| | Camera | Exposure | ISO | Colored Light | White Light | Filter |
|---|---|---|---|---|---|---|
| 1 | PhaseO | 1/8 | 400 | Royal Blue 450nm | NONE | NONE |
| 2 | PhaseO | 1/10 | 400 | Blue 465nm | NONE | NONE |
| 3 | PhaseO | 1/20 | 400 | Cyan 505nm | NONE | NONE |
| 4 | PhaseO | 1/15 | 400 | Green 535nm | NONE | NONE |
| 5 | PhaseO | 1/6 | 400 | Amber 570nm | NONE | NONE |
| 6 | PhaseO | 1/30 | 400 | Red 625nm | NONE | NONE |
| 7 | PhaseO | 1/6 | 400 | IR 700nm | NONE | NONE |
| 8 | PhaseO | 1/5 | 400 | IR 780nm | NONE | NONE |
| 9 | PhaseO | 1/2 | 400 | IR 870nm | NONE | NONE |
| 10 | PhaseO | 3s | 400 | IR 940nm | NONE | NONE |
| 11 | PhaseO | 1m00 | 800 | UV 365nm | NONE | UV-PASS |
| 12 | PhaseO | 10s | 800 | UV 365nm | NONE | UV-BLOCK |
| 13 | PhaseO | 1/15 | 400 | NONE | Panels | NONE |
| 14 | Nikon C | 2s | 800 | UV 365nm | NONE | NONE |
| 15 | Nikon C | 1/10 | 800 | NONE | Panels | NONE |

Figure 3.3: Graphical user interface of the acquisition system: (1) definition of object name and storage location; (2): manual control of MS lighting and filter wheel; (3): manual control of cameras; (4): user-defined shot list for MS imaging, which can be executed automatically; (5) controls for focus shift correction (see Section 3.2); (6) PS controls; (7) display of last recorded image, zoom-able and with histogram shown below.

on the ruler corresponds to a difference in depth of $l/\sqrt{2}$. To measure a focus shift with respect to a reference wavelength $\lambda_0$, the camera is first focused on the reference plane under $\lambda_0$. Subsequently, an image is taken under another wavelength $\lambda_i$. The focus shift for $\lambda_i$ in real-world units can be directly read by finding the sharpest area of the ruler in the corresponding image.

In this way, a series of measurements is performed: For each wavelength $\lambda_i$ relevant to the respective imaging system, with $1 \leq i \leq M$, and for several distances $d_j$ within the typical working range, with $1 \leq j \leq N$, where $N \geq 3$, an image $I_{i,j}$ is taken and used to measure the focus shift $\hat{\Delta}_{i,j}$. This procedure is summarized in Algorithm 3.1.

In a typical MS imaging setting, the utilized wavelengths are limited to a discrete set (defined by an available set of optical filters [Cos15] and/or narrow-band light sources [Kno18]), but the imaging distance can be varied continuously according to the individual use-case. It is therefore practical to consider individual one-dimensional functions $\Delta_i(d)$ per wavelengths $\lambda_i$, instead of the joint two-dimensional function $\Delta(\lambda, d)$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

**Algorithm 3.1:** *Outline of the calibration procedure.*

```
Data: wavelengths $\lambda_i$, sample distances $d_j$,
Result: samples $\hat{\Delta}_{i,j}$ of the unknown function $\Delta(d, \lambda)$
for $j \in [1, N]$ do
    place target plane at distance $d_j$ from the camera;
    focus on target plane under reference wavelength $\lambda_0$;
    for $i \in [1, M]$ do
        take image $I_{i,j}$ under wavelength $\lambda_i$;
        $\hat{\Delta}_{i,j} \leftarrow$ focus shift read from ruler in $I_{i,j}$;
    end
end
```

We found that $\Delta_i(d)$ are well approximated by quadratic polynomials of the form $\Delta_i(d|\boldsymbol{\theta_i}) = \boldsymbol{\theta_i}\,[d^2\; d\; 1]^T$. The parameter vector $\boldsymbol{\theta_i} \in \mathbb{R}^3$ is found via least-squares fitting to the measured data by minimizing:

$$\min_{\boldsymbol{\theta_i}} \sum_{j=1}^{N} (\hat{\Delta}_{i,j} - \Delta_i(d_j|\boldsymbol{\theta_i}))^2.$$

Under knowledge of the object distance, the thus-found functions are used to compute the magnitude of the relative focus shift $\Delta$ for a given imaging configuration. This focus shift describes the distance the focus plane moves in real-world units; it can thus be mechanically compensated by either adjusting the distance between camera and object or the focal distance of the lens by $-\Delta$.

### 3.2.2 Experimental results

The calibration approach was applied to the MS imaging system described in Section 3.1. For the reference focus on the target plane, a single wavelength from the center of the visible spectrum (535 nm) was chosen as $\lambda_0$. Five wavelengths from the UV (365 nm) and IR (700 nm, 780 nm, 870 nm and 940 nm) ranges were known to produce noticeable focus shifts with the camera system and thus considered for calibration. Distance samples were taken in 100 mm steps from 400 mm to 1000 mm, and in 200 mm steps from 1000 mm to 1800 mm. The aperture was set to f/4 in order to minimize the depth of field and thus ease the identification of the sharpest area on the ruler. As a calibration object, the commercially available *Spyder LensCal*, that is originally designed for autofocus calibration, was used. Figure 3.4 shows the full setup during calibration.

In this experiment, the focus shift magnitudes were read from the images of the tilted ruler manually. Figure 3.5a shows the measurements obtained and the quadratic polynomials fitted. Note that above 1200 mm working distance, some measurements are missing because the focus plane shifted beyond the depth region covered by the ruler plane.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

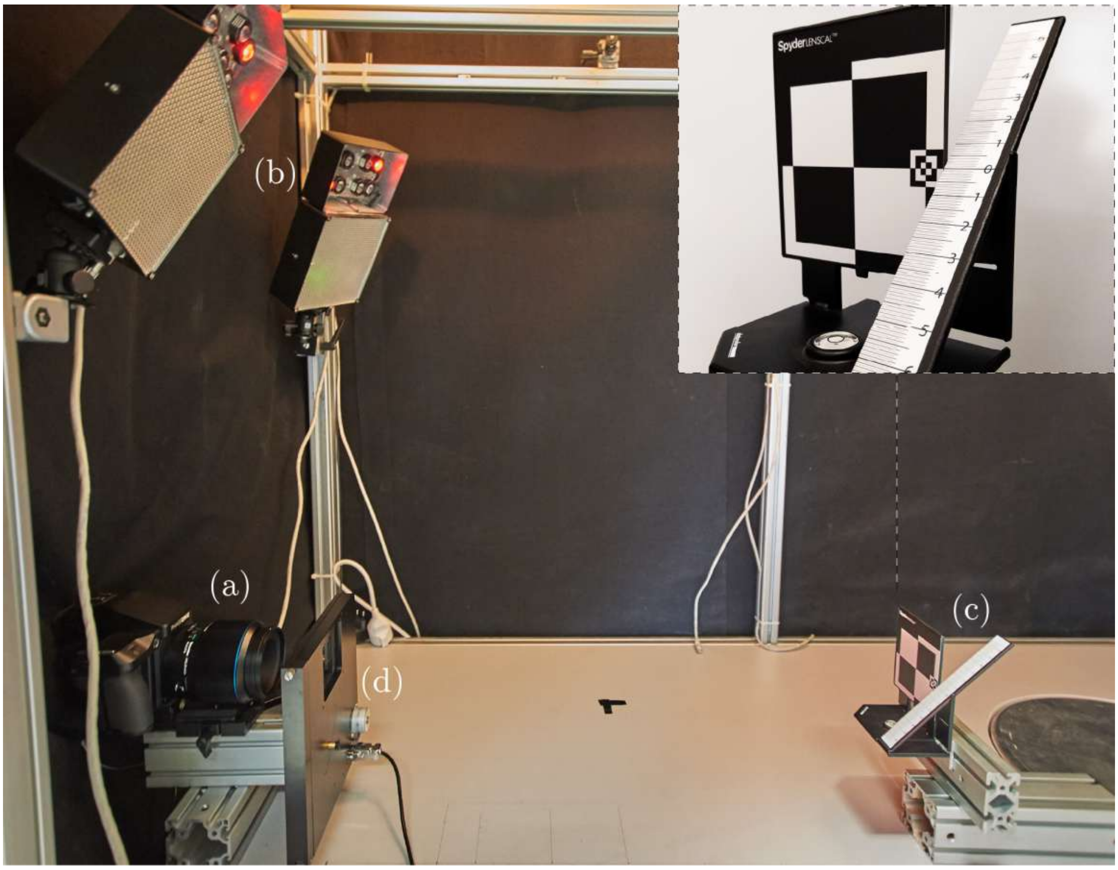


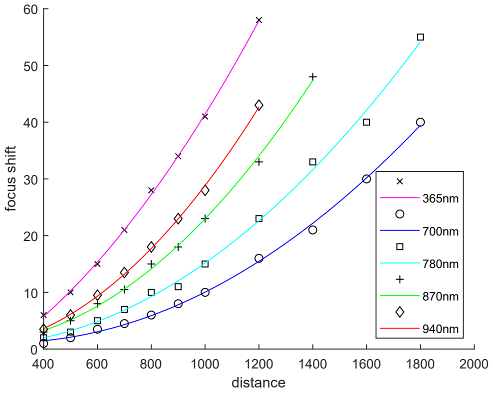
Figure 3.4: Calibration setup consisting of the camera (a), MS light sources (b) and the calibration object (c). An optical filter wheel (d) is necessary to block fluorescent light when imaging in the UV range.

The Root Mean Square Errors (RMSE) and coefficients of determination of the fitted functions with respect to the measurements are listed in Figure 3.5b.

To use the results for focus shift compensation in practice, the lens is first focused on the target object under $\lambda_0$. For any wavelength included in the calibration, the focus is adjusted according to the respective estimated function by driving the auto-focus motor via the camera API.[2] As shown qualitatively in Figure 3.6, the implementation of the proposed calibration and subsequent compensation into our system allows the acquisition of in-focus images in the UV and IR range, even though the camera lens used is not optimized for that purpose.

[2] Compensating the focus shift via the autofocus motor might require further calibration efforts, in order to find the relationships between motor steps and shifts of the focal plane. An alternative approach (conceptually simpler but more hardware-intensive) would consist in the adjustment of the physical distance between object and camera using a linear positioning unit.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



(a)

| $\lambda$ (nm) | RMSE (mm) | R$^2$ |
|---|---|---|
| **365** | 0.4522 | 0.9995 |
| **700** | 0.5616 | 0.9984 |
| **780** | 1.0608 | 0.9970 |
| **870** | 0.6531 | 0.9985 |
| **940** | 0.4607 | 0.9991 |

(b)

Figure 3.5: Experimentally determined compensation functions: (a) plots of measurements and fitted functions; (b) root mean squared errors and coefficients of determination (R$^2$) of the fitted functions.

### 3.2.3 Evaluation

To test the feasibility of a quadratic polynomial for modeling the focus shift with respect to object distance, an exhaustive cross-validation was performed: Polynomials of degrees 1-5 were fitted to different subsets of the distance samples and tested on the remaining ones. The values shown in the Table 3.3 aggregate results of all wavelengths and possible selections of samples. It can be observed that for more than three training samples, the 2$^{nd}$ degree polynomial best describes the data, while the other tested models over- or underfit. The table also shows the expected errors when calibrating the system with a reduced number of distance samples.

The focus correction method added to the MS image acquisition system and thereafter used in on-site imaging campaigns. Figure 3.7 shows a comparison of imagery acquired before and after the correction was implemented.

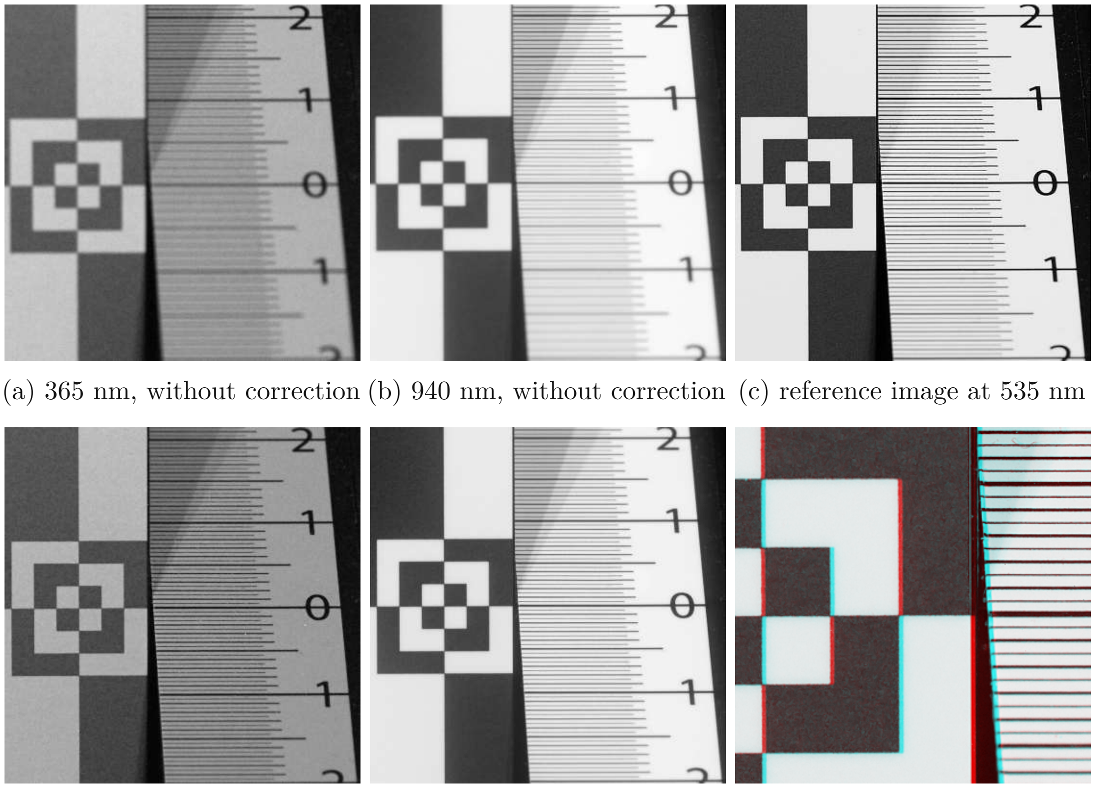
### 3.2.4 Discussion

As shown above, the proposed calibration and correction approach is suitable to enable the acquisition of sharp images outside the visible spectrum under usage of imperfect lenses. The model errors shown in Figure 3.5b lie in the order of 1 mm or below; this is well below the magnitude of depth variations that are to be expected in the carrier surfaces of graphical heritage (*e.g.*, historical manuscripts, paintings).

However, the presented implementation exhibits shortcomings, one of which is the amount of manual intervention required in the calibration procedure. During image acquisition, the distance between camera and calibration object must be adjusted for the sampling of different focal distances. This was done manually in the experiments described, but

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek, WIEN Your knowledge hub



(a) 365 nm, without correction (b) 940 nm, without correction (c) reference image at 535 nm

(d) 365 nm, corrected (e) 940 nm, corrected (f) 535 nm + 940 nm corrected

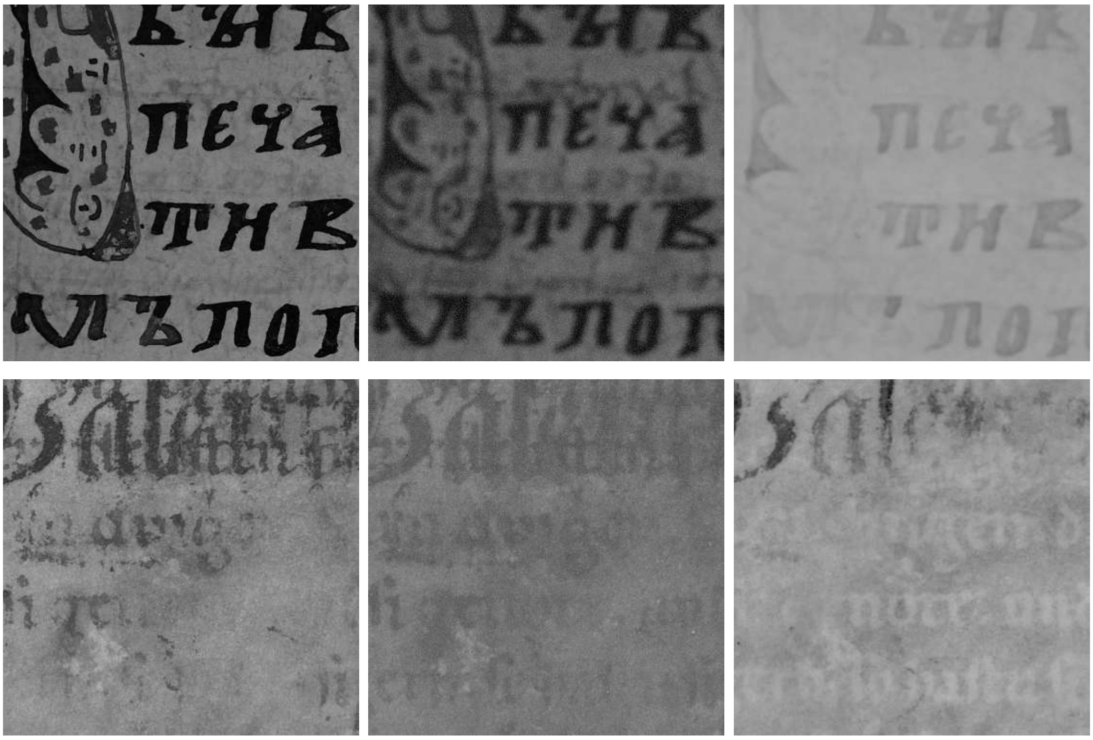
Figure 3.6: Images of the calibration object acquired with the MS imaging system, taken from 1 m distance and aperture f/8. The top row shows uncorrected images taken under UV (365nm) and IR (940nm) illumination, and the 535nm reference image. The bottom row shows the corrected UV and IR images as well as a composite of the 535 nm image and the corrected 940nm image (magnified); the misalignment introduced by re-focusing and the concomitant change of scale is visible in the form of color fringing.

the automation using a linear positioning unit is straightforward. Furthermore, the reading of focus shifts from the acquired images was done manually in the experiments described in this section, which is a time-consuming task when performed by a human. In later tests, the process was improved by automatically detecting the sharpest area on the ruler in terms of maximum second derivatives; the results thus obtained closely resembled the manual readings. A further, yet not implemented improvement would consist in automatically detecting the ruler and reading out the values corresponding to the sharpest areas; this could be supported by a modified calibration object, *e.g.* by adding uniquely identifiable markers [Fia05] to the ruler plane. Depending on the requirements of the application, the calibration process can be accelerated further by reducing the number of distance samples (see Table 3.3).

It is worth mentioning that while using the maximum aperture size is recommended for

| $N$ | *degree 1* | *degree 2* | *degree 3* | *degree 4* | *degree 5* |
|---|---|---|---|---|---|
| 2 | 11.37 | | | | |
| 3 | **9.01** | 12.20 | | | |
| 4 | 8.26 | **5.16** | 85.84 | | |
| 5 | 7.06 | **4.66** | 8.58 | 402.39 | |
| 6 | 6.79 | **3.55** | 31.68 | 157.12 | 1778.73 |
| 7 | 6.04 | **1.83** | 2.49 | 80.81 | 688.40 |
| 8 | 4.60 | **0.74** | 3.37 | 13.61 | 53.53 |
| 9 | 3.54 | **1.02** | 1.78 | 4.15 | 13.05 |
| 10 | 2.99 | **0.70** | 0.99 | 2.05 | 4.10 |

Table 3.3: Results (RMSE) of exhaustive cross-validations for different choices of polynomials. $N$ measurements were used for training, $11 - N$ for testing. The lowest error for each value of $N$ is bold.



(a) 535 nm (b) 365 nm (c) 940 nm

Figure 3.7: Top row: images taken with the MS imaging system before focus correction was implemented (Slepče Apostolus, Ivan Vazov Library, Plovdiv). Bottom row: images taken with the focus correction implemented (Fragment no. 6, Maria Taferl). Images show 3 cm × 3 cm sections of medieval manuscripts, shot with an f/8 aperture.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

the calibration process, a smaller aperture may be appropriate in production imaging to ensure that larger parts of the object imaged are in focus. While in an ideal lens, the aperture size should not move the focal plane, this must not necessarily hold for real-world optical blocks, such that this aspect should be verified when calibrating with a different aperture than used for productive imaging.

Finally, a crucial issue to consider is the change in image scale when adjusting the camera-object distance or the focal distance, as this introduces misalignments between spectral layers (see Figure 3.6f for illustration). However, misalignments can also occur for other reasons, such as lateral chromatic aberrations or the application of optical filters. Hence, fine-registration is a sensible processing step of any MS image processing pipeline, regardless if an explicit focus correction takes place or not.[3] Dedicated methods for the registration of MS images of graphical heritage are found in literature - see, for example, Lettner *et al.* [LDSM08] or Jones *et al.* [JCT+19]. Alternatively, borrowing methods from medical imaging can be considered. For example, Heinrich *et al.* propose the Modality Independent Neighborhood Descriptor (MIND) [HJB+12] for the registration of multi-modal medical imagery, *e.g.*, computed tomography to magnetic resonance imaging. Based on image self-similarity, the dependence on the specific gray value distributions in the respective modalities is reduced - a property is also relevant for MS imaging, where different gray value distributions must be expected in different spectral ranges. When used with a deformable registration framework, this approach can accommodate deformations of organs that must be expected in medical imaging [HJB+12], but also minor deformations in graphical heritage objects which can be expected, *e.g.*, in manuscript leaves. Figure 3.8 shows an example where the 700 nm and 365 nm layers of a medieval manuscript were fine-registered with MIND and the originally proposed registration framework [HJB+12].

Considering the complications introduced by the proposed calibration and correction approach, its usefulness for practical applications must be discussed. Arguably, the most convenient and reliable solution for a user would be the use of a lens corrected for all wavelengths of interest. While perfect color correction for an arbitrary number of wavebands within the UV, visible and IR range cannot be achieved, several authors report satisfying results in MS imaging of graphical heritage, using corrected lenses as the only measure against chromatic aberration [MJH19, EKC+10, KMB22]. In other cases, lenses specified as color-corrected by the manufacturer introduce intolerable focus shifts [KCM21].[4] I thus argue that the benefit of introducing a dynamic correction approach depends on the chromatic aberration characteristics of the lens used, as well

[3] Pixel-accurate registration (*i.e.*, that in each layer, a given pixel corresponds to the same imaged surface point) is a prerequisite for any use of MS images where multiple layers are combined - be it simply the creation of a false color image, or any statistical method where each pixel of an n-layer MS image is treated as an n-dimensional observation of a surface point.

[4] In own tests with the *Schneider Kreuznach Citrine* VIS-IR lens used by Kleynhans *et al.* [KCM21], allegedly corrected from 400 nm to 700 nm, intolerable focus shifts were measured. For example, when focusing on a target at 50 cm distance under 535 nm illumination, the focus approximately shifts by 2.5 cm at 450 nm, by 2 cm at 700 nm and by 6 cm at 940 nm.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

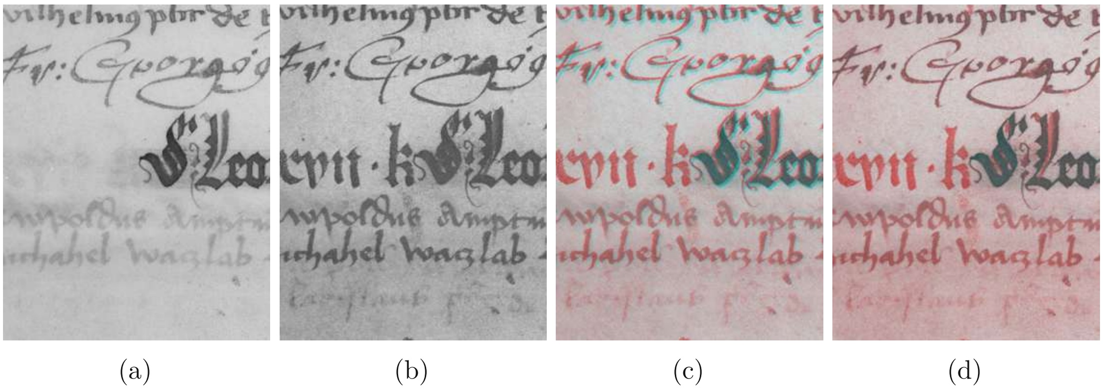


Figure 3.8: Registration with the MIND descriptor: (a) 700 nm layer; (b) 365 nm layer; (c) false color image with 700 nm in red channel and 365 nm in green and blue channels - note the presence of color fringing; (d) same false color images, with layers registered - note the absence of color fringing.

as the application-specific requirements on image sharpness. Compared to manually adjusting the focus [Kög14, DVC13] or auto-focusing between shots [BCLP98, PMd23], the presented approach ensures a fast and consistent acquisition process without introducing additional optical elements [RL15].

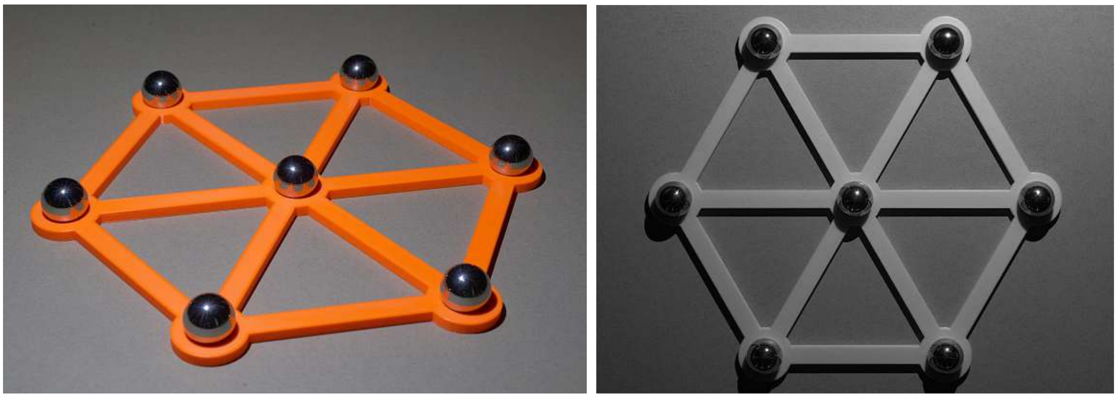
## 3.3 Estimating light positions for photometric stereo

In the following, a practical heuristic for the calibration of PS light source locations is described. The main purpose of this section is to aid the documentation and repeatability of experiments described later in this work (where this heuristic is applied) and not to introduce or evaluate any novelties themselves. On the other hand, readers concerned with practical imaging might find the descriptions useful for their own work.

As the imaging system described in Section 3.1 is modular and the locations of light sources with respect to the camera are not fixed, they must be calibrated for each specific setup (if their positions are required for the reconstruction method used). Thus, an efficient and convenient calibration scheme is desirable. The approach used in this work is based on images of specular spheres, coarsely following the method of Powell *et al.* [PSG01].

For ease of use, a custom 3D-printed mount for holding the specular spheres was designed. As shown in Figure 3.9a, it holds seven steel balls of 20mm diameter in a regular triangular grid with 100mm edge length. For each light source position to be calibrated, an image of the calibration object (*i.e.*, mount with steel balls) is acquired. Depending on the specific acquisition setup, not the whole calibration object might fit into the field of view (as in the example in Figure 3.10), but the method works as long as two balls are visible. For the estimation of light positions from the calibration images acquired, three sub-problems

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b)

Figure 3.9: A 3D-printed mount for seven steel balls, used as a light source calibration object: (a) color image from an oblique angle and (b) example of a calibration image acquired with the PS acquisition system.

must be solved:

1. Detection of spheres
2. Detection of highlights on the spheres
3. Triangulation of light positions

In the following, the specific implementations for each of these steps that are used throughout this work are elaborated.

### 3.3.1 Detection of spheres

The specular spheres reflect their surroundings and, due to the specific illumination conditions, cast hard shadows; these additional circle-like shapes appearing in the calibration images (see Figure 3.9b) make the precise detection of the spheres non-trivial (cf. Section 2.2.4). In the following, an alternative to the heuristic proposed by Liao *et al.* [LBT$^+$17] is presented, that exploits the geometry of the mount holding the reflective balls.

Initial trials with a mount made of black plastic revealed that the spheres appear mostly black when illuminated with a single point light source (see Figure 3.9b for an example) and thus cannot be discriminated from the background. The material of the mount was thus changed to a bright orange, but still, problems are caused by hard shadows cast by the spheres. Thus, the following heuristic is developed, making use of geometrical properties of the calibration object.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

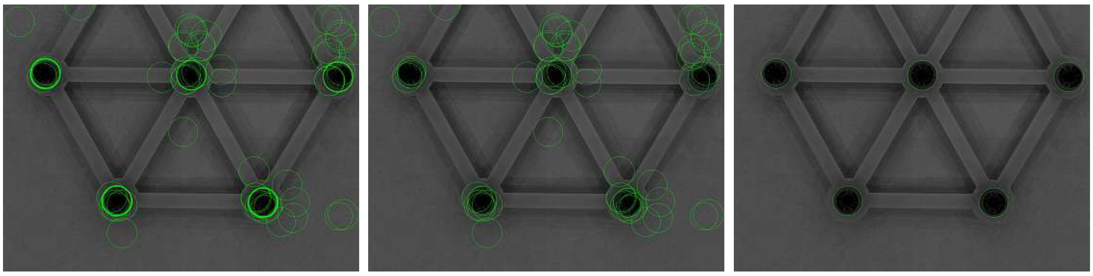


Figure 3.10: Detection of balls, visualized as green circles on the mean image of all input images used: (a) raw Hough-transform detections from multiple input images; (b) similar detections are clustered; (c) filtered detections, using Algorithm 3.2.

1. **Circle detection.** Circles are detected in input images via Hough transform. Knowing the ground sampling distance and the physical size of the balls, detected circles that deviate more than ±5% from the expected radius in image space are discarded. Experimentally it was determined that input images with low incident illumination angles are most suitable for the purpose, such that by default, the lights at 13° and 32° of the acquisition setup are used. While generally, spheres are mapped to ellipses under perspective projection [LBT+17], the deviation from circles was not noticeable with the imaging setup at hand.

2. **Merging circles.** Circles obtained from different input images are clustered if both the distance between their centers and radii deviate by less than 5% of the expected radius. A cluster is then represented by a single circle with median radius and center position; also, the number of the circles forming the cluster is stored.

3. **Filtering circles.** To remove false detections, the regular triangular arrangement of balls is exploited. The concocted filtering algorithm is based on the insight that for each node in a regular triangular grid, the distance to the nearest neighbor must be exactly the edge length of the grid - detected circles that violate this condition are successively removed, following the strategy outlined in Algorithm 3.2 (where the cluster size obtained in the previous step serves as "detection confidence").

The detection scheme outlined above is a heuristic and cannot be guaranteed to deliver the correct balls. Especially Algorithm 3.2 is not guaranteed to converge to the right result - failing examples can easily be constructed, *e.g.*, if all detection confidences (cluster sizes) are equal, or if wrong detections are also arranged in a rectangular grid of the same edge length. The only evaluative statement that can be made at this point is that within the 11 acquisition setups calibrated with the method (a total of 67 balls, 4-7 visible per setup), 100% of balls were detected pixel-perfectly (by visual inspection) and no false detections were made.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

**Algorithm 3.2:** Filtering detected objects that should be arranged in a regular triangular grid.

```
Input: a list $Q$ of objects $q_i$, with positions $q_i.p$ and detection confidences $q_i.c$;
       expected edge length $d$ of the regular triangular grid; distance tolerance $\epsilon$
Output: A list $Q' \subseteq Q$ of objects that lie on a triangular grid with given edge
        length.
$regular \leftarrow False$;
$Q' \leftarrow Q$;
while $\neg regular$ do
    $R \leftarrow \emptyset$ ;                                /* objects to remove */
    foreach $q_i \in Q'$ do
        $q_j \leftarrow nn(q_i)$ ;      /* nearest neighbor w.r.t.  position */
        $d_{ij} \leftarrow |q_i.p - q_j.p|$;
        if $d_{ij} > d + \epsilon$ then
            /* Too far away from any other object - can be
               safely removed.                                          */
            $R \leftarrow R \cup q_i$;
        end
        else if $d_{ij} < d - \epsilon$ then
            /* Two objects are too close - remove the one with
               lower detection confidence or both, if they are
               equal.  This assumes that correct objects
               always have a higher detection confidence than
               conflicting incorrect objects.                    */
            if $q_i.c \geq q_j.c$ then
                $R \leftarrow R \cup q_j$;
            end
            if $q_i.c \leq q_j.c$ then
                $R \leftarrow R \cup q_i$;
            end
        end
        $regular \leftarrow R = \emptyset$ ;  /* If no more irregularities are found,
         we are done.  */
        $Q' \leftarrow Q' \setminus R$ ;                  /* update result list */
    end
end
return $Q'$;
```

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

### 3.3.2 Detection of highlights

Given that the currently active LED is the only light source in the imaging environment, a straightforward solution would be to detect the single brightest blob within each sphere. However, in some cases inter-reflections from the ball mount also cause specular highlights of similar brightness (like in Figure 3.11a).

Assuming that no ball is illuminated from an angle greater than 90° from the viewing direction (in reasonably designed PS acquisition setups, a plausible assumption [DC05]), the point of reflection cannot occur at an angle greater than 45° from the viewing direction (relative to the sphere center); thus, the search radius in image space can be reduced to $sin(\frac{\pi}{4})r$, where $r$ is the radius of the detected circle. This effectively excludes reflections stemming from the ball mount, which arrive at the sphere at a larger angle and thus lie outside the search radius. Blobs are detected as extrema in a Laplacian-of-Gaussian cube, and the blob with maximum mean intensity is chosen. See Figure 3.11a for illustration.

Like the ball detection algorithm, no formal evaluation is carried out; only by visual assessment, a success rate in 100% of tested cases (in total, 2562 detected highlights from 11 calibrated setups) can be reported.

### 3.3.3 Triangulation of light positions

Following Powell *et al.* [PSG01], positions of ball centers and highlights are determined in object space, and then imaginary rays emerging from the camera and reflected at highlights are computed, under knowledge of internal camera parameters and physical sizes of spheres. In contrast to Powell *et al.* [PSG01], more than two spheres are supported, such that the closest point to multiple rays (*e.g.*, an approximate intersection) must be found; this is solved with a least squares approach [Tra13]. The principle is illustrated in Figure 3.11b. It is worth noting that correcting for geometrical lens distortions is generally advisable as a pre-processing step to the calibration procedure described. However, due to the mild distortion characteristics of the lens used (see Section 3.1), this step was omitted in this work.

## 3.4 Summary

The mobile acquisition system for MS imaging and PS used for experiments in the remainder of this work consists of an achromatic medium format camera, a filter wheel, MS LED panels and a dome-shaped arrangement of white LEDs. All components are integrated with custom software to enable automated acquisition processes. Longitudinal chromatic aberrations in MS images are prevented with a novel approach for calibration and mechanical correction at acquisition time. Light source positions for PS are calibrated with a traditional approach based on reflective spheres, whereby the sub-problem of precisely segmenting the spheres is solved using their regular arrangement.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

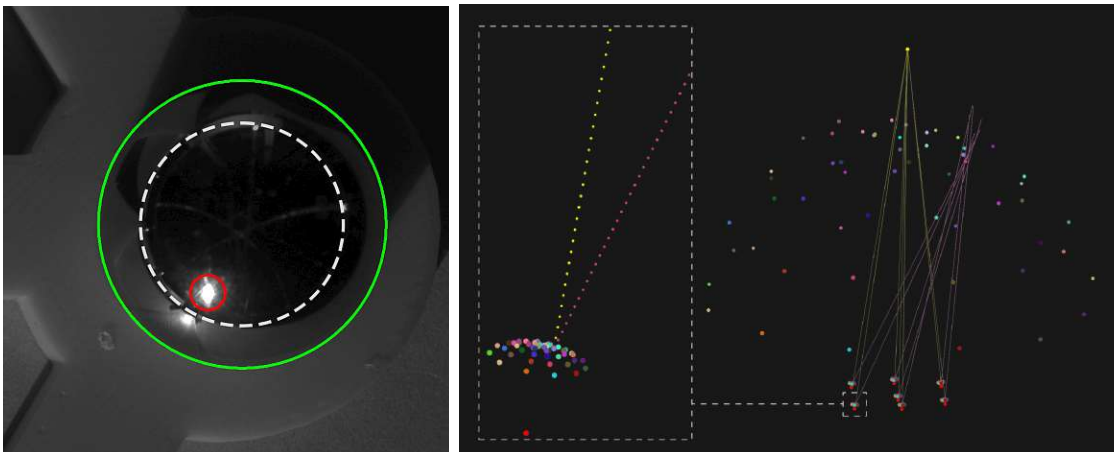


Figure 3.11: (a) highlight detection: the steel ball is outlined in green and the detected highlight in red. The white dashed circle with radius $sin(\frac{\pi}{4})r$ (with $r$ being the radius of the green circle) demarcates plausible reflections of light sources from inter-reflections from the mount. (b) triangulation of light sources: one set of rays emerging from the camera center (topmost point), reflected at the spheres (sphere centers in red), and approximately meeting at the light source position. Other colored dots show the light source positions and reflection points on spheres.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

# CHAPTER 4

# Photometric stereo

Near, far, wherever you are - I
believe that the heart does go on.

*Celine Dion*

This chapter is concerned with aspects of Photometric Stereo (PS) relevant for the recording of graphical heritage manifesting in depth variations relative to a carrier surface. One part of the chapter investigates the impact of the number and placement of light sources on the reconstruction of approximately planar objects. The next (and longer) part goes into depth about the errors introduced by assuming a parallel lighting model in a point light acquisition setup; an associated error mitigation strategy is especially useful for cases where the microstructure of an object is of greater interest than its macrostructure. At the end of the chapter, the merits of PS and the insights gained in this work are demonstrated in a case study, where ink-less typewriter impressions in paper are recorded and visualized in order to open new data sources in literary studies. First of all, however, the datasets used for experiments in the subsequent sections are described.

## 4.1 Datasets

Two PS datasets, both acquired with the imaging setup described in Section 3.1, are used for experiments in the remainder of this chapter. The *Roman coins* dataset comprises 22 image sets but lacks light source calibration and a ground truth. The *Diffuse objects* dataset contains only four regular image sets plus one image set for light source calibration and features ground truth surface normals. Images of both dataset are not corrected for lens distortions.

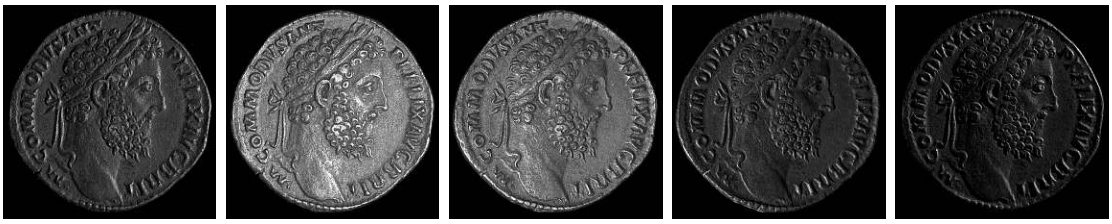


Figure 4.1: Example input images of the *Roman coins* dataset.

### 4.1.1 Romain coins

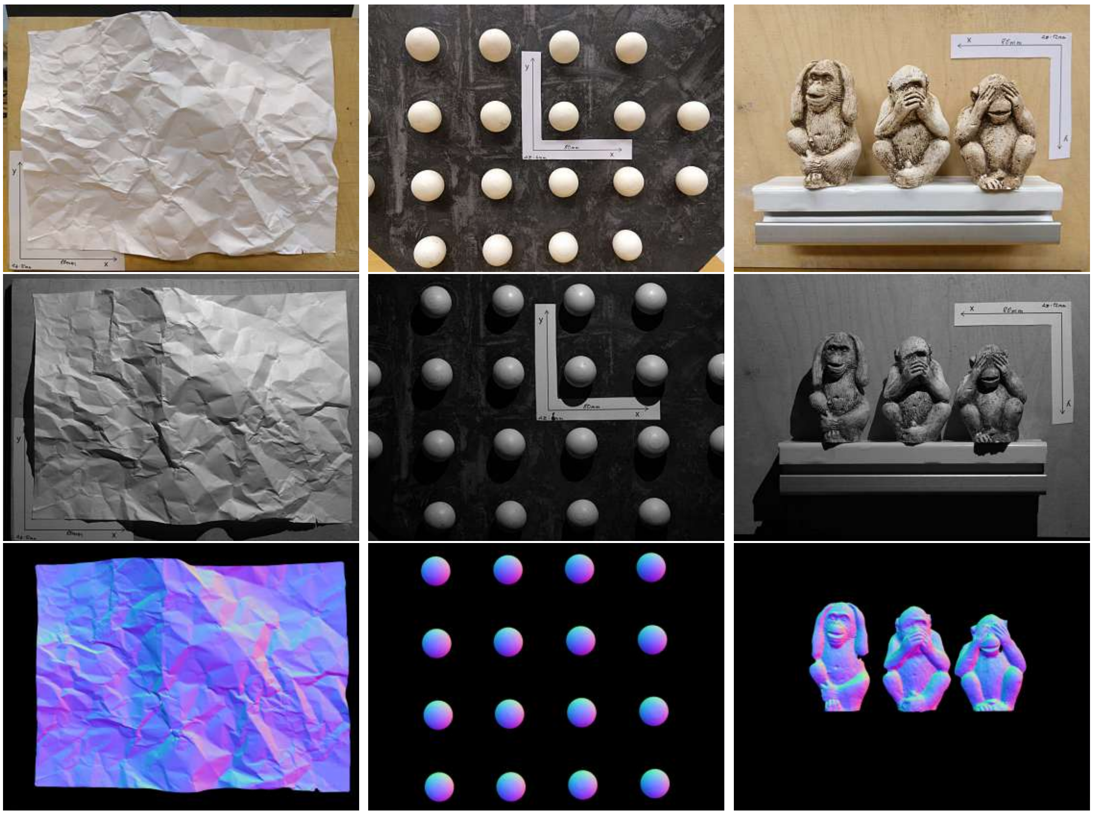
The test objects are 11 ancient coins from the Roman Empire, made from brass, copper, silver, gold or billon and showing various degrees of corrosion. Their diameters range between 19 and 33 mm. Each coin is imaged from both sides, resulting in 22 image sets. For imaging, the coins are placed approximately in the center of the camera frame. For each light source available with the acquisition setup (see Section 3.1), an image is taken from a vertical distance of ca. 600 mm. With that setup and a focal length of 120mm, the object surface was sampled with approximately 45 pixels per mm. Masks for the coins are created manually in *Adobe Photoshop* and all images are cropped to the size of the respective coins; see Figure 4.1 for examples of masked and cropped input images. For this dataset, neither light source calibration information nor ground truth surface normals are available.

### 4.1.2 Diffuse objects

Four test objects with mostly diffuse surface characteristics were crafted for the purpose of this dataset: a flat gray cardboard as a trivial case (*'flatfield'*), a wrinkled sheet of white paper (*'paper'*), a set of wooden spheres painted matte white (*'whiteballs'*), and three wise monkeys made of a plaster-like material (*'monkeys'*) - see top row of Figure 4.2. Light source positions are estimated with the approach described in Section 3.3.

Ground truth surface normals are generated from 3D models acquired with an industrial structured light scanner (*AICON PrimeScan*, see Figure 4.3). The 3D models are aligned to the PS images using the *Raster alignment* tool of *MeshLab* [CCC$^+$08], in a similar way as described by the authors of the *DiLiGenT* [SWM$^+$16] and *LUCES* [MLBC21] datasets. Surface normal maps are generated directly in *MeshLab*, by exporting images rendered by a custom shader program; note that the alignment method and rendering in *MeshLab* assume a pinhole camera model, such that minor misalignments with respect to the undistorted input images are expected. Binary masks segmenting objects from background are created manually in *Adobe Photoshop*. In Figure 4.2, gray-scale images, masked ground truth normal maps and physical dimensions of the test objects are shown.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) paper
255x185x49mm

(b) whiteballs
195x195x28mm

(c) monkeys
166x87x64mm

Figure 4.2: *Diffuse objects* dataset. From top to bottom: color image of the test object, example input image, and visualizations of ground truth normal maps with object masks applied. The trivial *flatfield* object is omitted here.

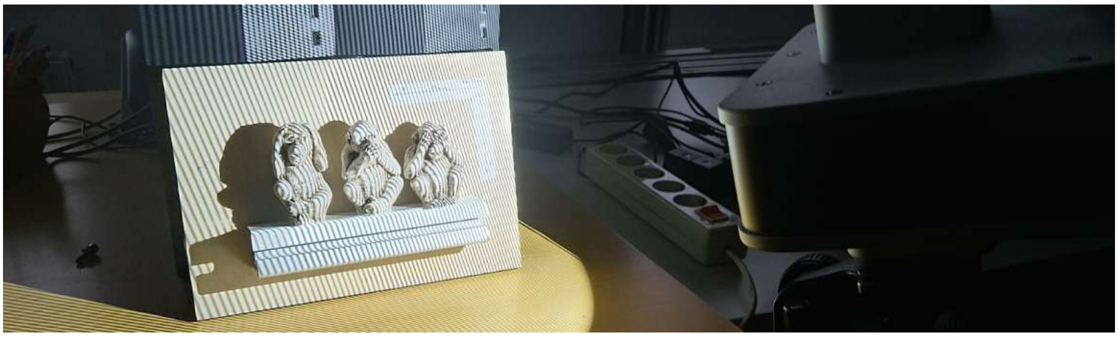


Figure 4.3: Generating ground truth for *Diffuse objects* dataset with a structured light scanner.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

## 4.2 Investigations on light source configurations

While more input images (*i.e.*, light sources) are expected to increase the robustness of PS to noise and model outliers [WGS+11], their acquisition also increases collection time and computational costs. In this section, the effects of different numbers and arrangements of light sources on PS reconstruction quality are investigated, in order to inform necessary trade-offs between costs and reconstruction quality. Experiments are described with both datasets described in Section 4.1. For the *Diffuse objects* dataset, surface normals are computed using a calibrated PS approach and evaluated against the included ground truth; for the *Roman coins* dataset, however, neither light source positions nor ground truth is available, such that surface normals are computed using an uncalibrated PS approach and the reconstruction using all available input images is used as a reference solution, against which the results of reduced configurations are compared. The results suggest that a dome setup with a large variation of light directions, as available in the acquisition setup described in Section 3.1, is unjustified for the reconstruction of graphical heritage; similar results are achievable with a fraction of well-placed light sources.

### 4.2.1 Experiments

In order to observe the influence of different numbers and arrangements of light sources on PS reconstruction quality, surface normals are computed using a variety of configurations and compared against ground truth normals. As an error metric, the Mean Angular Error (MAE) of the surface normals is used, which is a common practice in PS literature [AG15, QLD15]. The following classes of configurations are considered:

1. **Random selection:** to test the general assumption that reconstruction quality increases with an increasing number of light sources/input images, reconstructions are made with $n$ randomly selected light sources, with $n \in \{3, 8, ..., 53\}$.
2. **Single elevation:** to test how closely the performance of the full dome arrangement can be approximated with circular configurations, reconstructions are made with rings of light sources of a single elevation angle (see Figure 3.2c for light source arrangement).
3. **Double elevation:** the circular configurations at the two best elevation angles determined at the previous point are combined, to see if this leads to a further improvement.

The execution of experiments differs for the two datasets used, such that the details on surface normals reconstruction and evaluations are described individually:

#### Roman coins

The first set of experiments was conducted using the *Roman coins* dataset, with the aim to evaluate light source configurations for PS reconstruction of historical coins

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

and similar quasi-flat objects carrying graphical heritage [BZS18]. For this dataset, no images for light source calibration were made; thus, an uncalibrated PS approach must be used. Specifically, the approach by Quéau *et al.* is used [QLD15], where the generalized bas-relief ambiguity is resolved by minimizing total variation in the results. As suggested by the authors, specularities and shadows are removed from the input data using the pre-processing scheme proposed by Wu *et al.* [WGS$^+$11]. The method by Quéau *et al.* assumes orthographic projection and parallel lighting, which is a reasonable approximation given the size of the observed objects versus the distance of camera and light sources in the imaging setup (for an extensive discussion of this topic, the reader is referred to Section 4.3). As no external ground truth is available for the *Roman coins* dataset, the reconstruction using all 54 available input images is used as a reference solution, against which reconstructions with reduced light configurations are compared.

#### Diffuse objects

The second set of experiments was conducted at a later point in time, in order to test the applicability of previous results for different kinds of objects and PS reconstruction methods. For the *Diffuse objects* dataset, light position estimates are available, such that a calibrated PS approach can be used. As with this dataset, the objects are much larger in relation to the light sources, the patch-wise normals reconstruction approach proposed in Section 4.3.1 is used. Additionally, another set of reconstructions is computed with the pre-processing scheme by Wu *et al.* [WGS$^+$11] applied (in the following, denoted 'robust PS'). Estimated normals are directly compared to the structured light based ground truth. Only the three non-trivial objects (*paper*, *whiteballs* and *monkeys*) are used.

### 4.2.2 Results

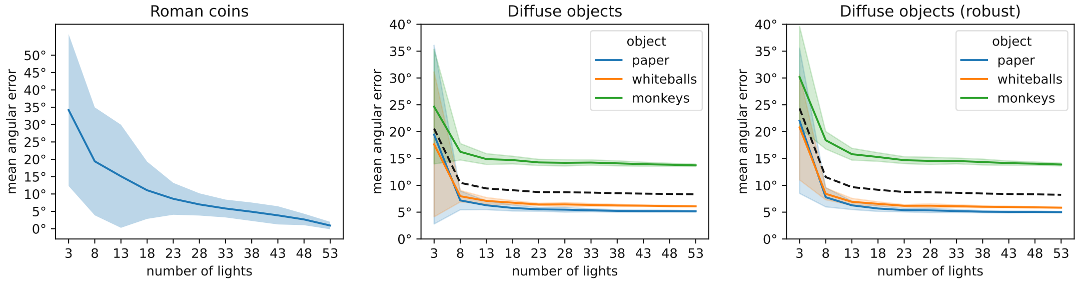
Results of the experiments are presented and interpreted in the following, grouped by the configuration classes defined at the beginning of the previous section.

#### Random selection

For the *Romain coins* dataset with 22 image sets, the reconstruction was repeated with three random light samples per object and per light count, leading to 66 evaluated reconstructions per light count. For the *Diffuse objects* dataset with three objects, 22 random light samples per object were evaluated, leading to the same total number of evaluations per light count.

Figure 4.4a plots the MAEs averaged across all measurements on the *Roman coins* experiments. Note that mean error and standard deviation both decrease with an increasing number of lights. This leads to the assumption that for constellations with few lights, their positioning has a large impact on the results. A similar behavior is observed in the results of the *Diffuse objects* experiments (Figure 4.4b). Here, the reconstructions are compared to external ground truth instead of the full dome reconstruction, such that the MAEs approach a residual error instead of zero. Because this residual error (resulting

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a)

(b)

(c)

Figure 4.4: Mean angular errors of reconstruction over the number of randomly selected light directions: (a) results for the *Romain coins* dataset, averaged over all image sets; (b,c) results for the *Diffuse objects* dataset, split by test object (dashed line shows mean across all objects). Error bands show the standard deviations.

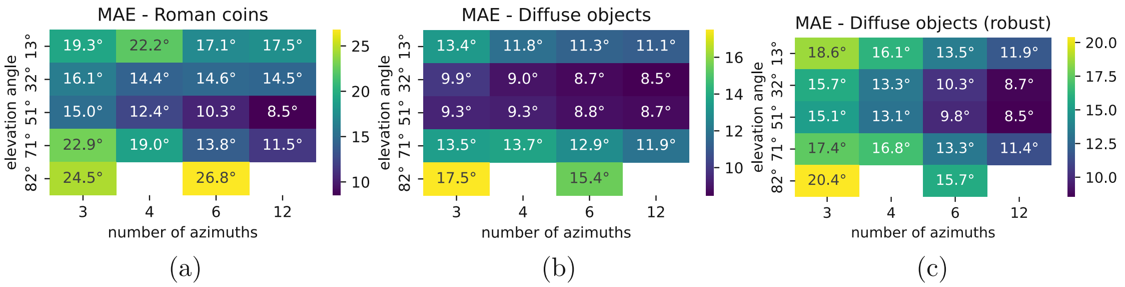
from inaccuracies in calibration, deviations from the assumed reflectance model, *etc.*) strongly depends on the test object, the corresponding results are plotted individually, such that the decreasing standard deviation within different random light samples can be observed. Using robust PS, the errors are larger for fewer lights but otherwise behave similarly. Motivated by these results, the following experiments aim at investigating optimal configurations for setups using a lower number of lights.

### Single elevation

In these experiments, 3, 4, 6 and 12 roughly equally spaced light positions on circles of equal elevation are considered. The goal is to evaluate the substitution of a light dome with a circular light arrangement, and which elevation angle would be preferable for such a setup.

Figure 4.5 shows the resulting mean errors for *Roman coins* and *Diffuse objects*. The light configuration with 82° elevation angle (close to the camera) shows the worst performance for both sets of experiments. This is explained by the ill-conditioned reconstruction problem for similar light directions [AG15]. On the other hand, very oblique lighting generally leads to more shadowed areas, negatively impacting quality [Woo80] - this can also be observed in the 13° results. For the *Roman coins*, the best results were achieved at an elevation angle of 51°. For the *Diffuse objects*, the situation is more ambiguous: errors are similar at 32° and 51° elevation angle, while classical PS performs slightly better on the former and robust PS performs slightly better on the latter. For all experiments, the errors tend to decrease with an increasing number of light directions per elevation, although there are some exceptions, especially with the *Roman coins*. In the *Diffuse objects* experiments, classical PS performs better with lower light counts than robust PS - a behavior also observed by Wu *et al.* [WGS$^+$11].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b) (c)

Figure 4.5: MAEs of circular light arrangements of constant elevation angle, for *Roman coins* and *Diffuse objects* (b,c). Side notes: color scales are different for each of the plots, as they should visualize the errors of different light configurations within each experimental setup, rather than between setups; 4 and 12 light configurations are missing at 82° because only 6 lights are available in the acquisition system at this angle.

### Dual elevation

The two best-performing elevations from the previous experiment (32° and 51°) are used to observe the effect of distributing a number of light directions across two circles of elevation instead of one. The same light azimuths as in the previous experiment are used, but on two elevation angles simultaneously. Additional tests were made for the three- and six-light settings, with the active lights on the upper circle rotated to a complementary configuration, as illustrated in Figure 4.6b. This experiment is only done with *Roman coins*.

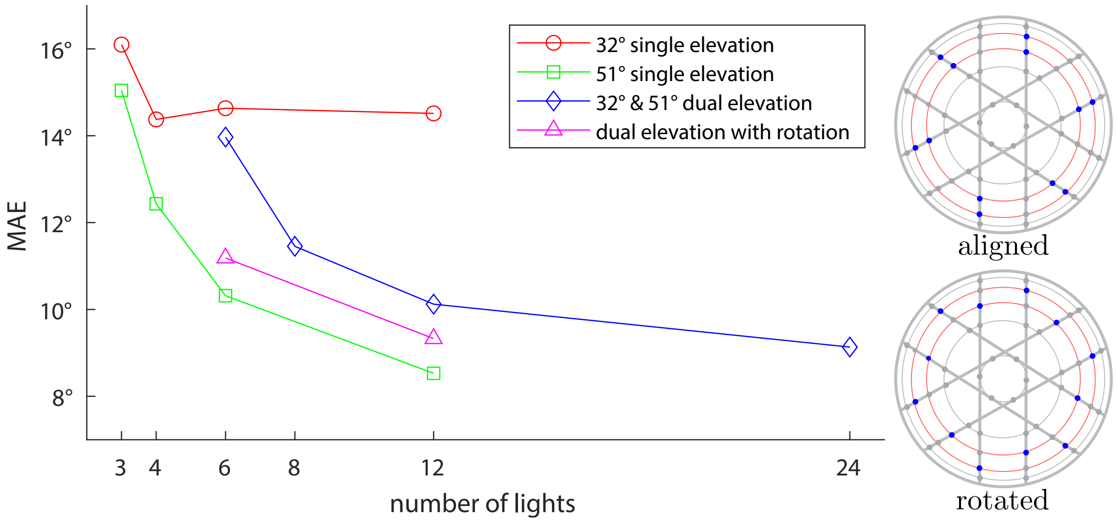
The results shown in Figure 4.6a can be interpreted as follows. Starting from the optimal single elevation of 51° (green), just doubling the lights on a less optimal elevation without adding a new azimuthal direction (blue), does not generally improve the outcome (compare, for example, the single configuration (green) with 12 lights and the double configuration (blue) with 24 lights). With a fixed total number of lights, placement only at the optimal single elevation is always superior. The rotated configuration (magenta) shows better results than the aligned configuration (blue), but the error still lies above the optimal single elevation with the same number of lights.

## 4.2.3 Discussion

The results of the single elevation experiments suggest that the optimal elevation angle of incident light lies between 32° and 51°. This can be related to previous results, where a slant angle of 54,74° (corresponding to an elevation of 35,26°) was determined as a theoretical optimum [DC05]. Given that for the experiments where the pre-processing method by Wu *et al.* [WGS$^+$11] was applied (*Roman coins* and *Diffuse objects* with robust PS), the lowest errors were measured at 51°, versus 32° for the remaining set, one could hypothesize that the pre-processing shifts the optimum to a greater angle.

Furthermore, the results of the dual elevation experiments suggest that when starting with a circular arrangement at such an optimal elevation angle, adding lights at new

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



(a) (b)

Figure 4.6: (a) Errors for two single elevations and their combinations, for *Roman coins*. The horizontal axis refers to the total number of active lights. (b) Illustration of the aligned (top) and rotated (bottom) version of a dual elevation setup. Active light positions are shown in blue.

azimuths is more beneficial than adding lights at new elevations.

Comparing the MAEs of random selections and single elevations, it is evident that for a given number of light sources, the expected errors for a circular arrangement at optimal elevation are generally lower than for a random distribution on a hemisphere. For the *Diffuse objects* experiments, MAEs over all datasets stay at approximately 10° between 13 and 53 light sources; the MAE for 12 light sources at optimal elevation is only 8.5°.

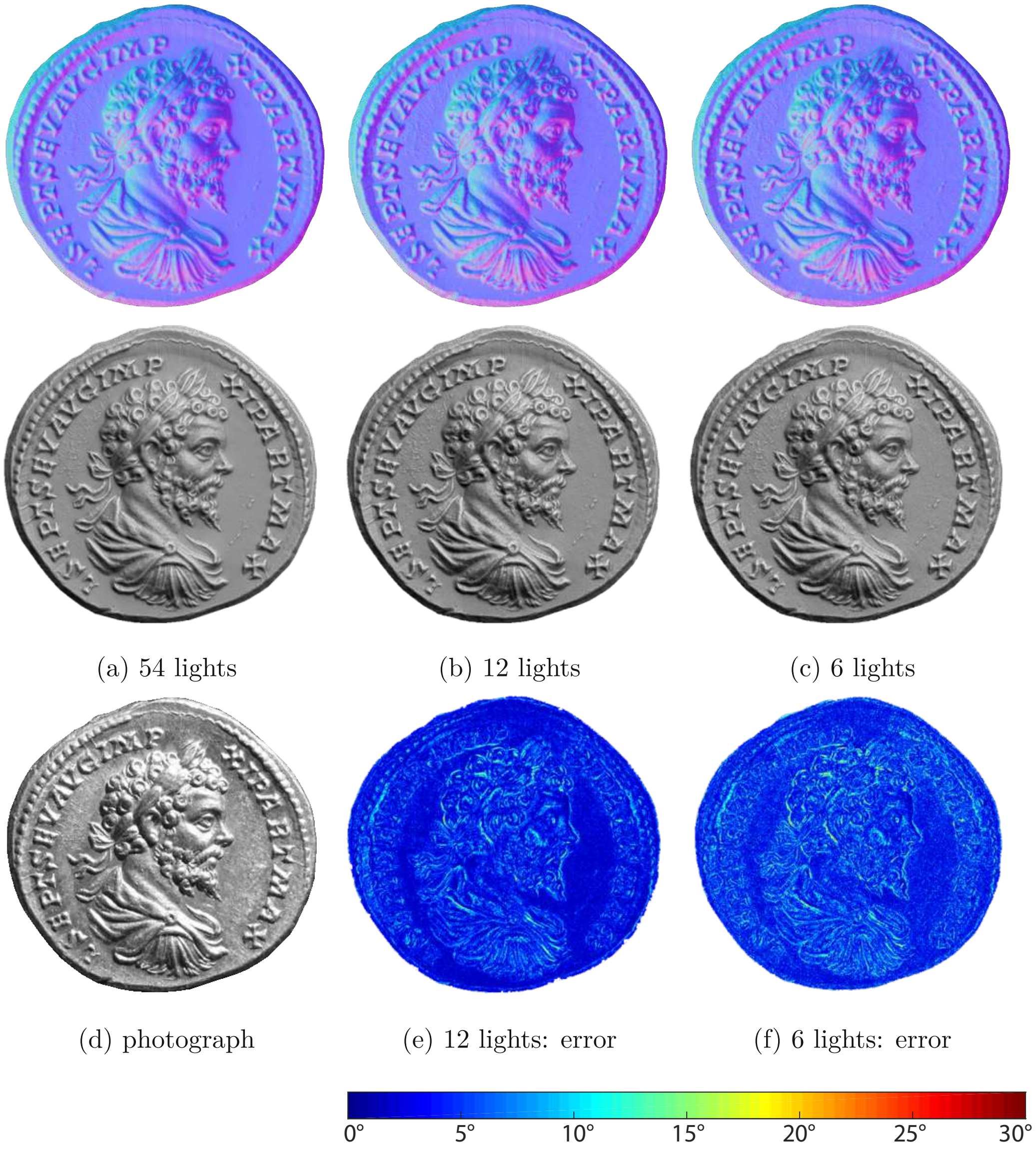
Thus, it can be concluded that for historical coins and similar objects carrying graphical heritage, the replacement of a light dome with a circular light arrangement is a reasonable option for downsizing imaging equipment and reducing reconstruction cost. Figure 5.4 shows a qualitative result. The reconstructions from 12 and 6 light sources at 51° elevation are visually not distinguishable from the full-dome reconstruction and high error values appear only locally at edges and cavities.

## 4.3 Classical Photometric Stereo in Point Lighting Environments

In the original formulation of PS [Woo80], parallel lighting (*e.g.*, by infinitely distant light sources) is assumed; thus, the light direction and intensity is considered equal for each surface point imaged, regardless of its location in space. This enables a straightforward estimation of surface orientations [Woo80] and the computation of depth in an independent

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) 54 lights

(b) 12 lights

(c) 6 lights

(d) photograph

(e) 12 lights: error

(f) 6 lights: error

Figure 4.7: Example reconstruction of a Roman gold coin. Top two rows: normal maps and shaded renderings of reconstruction results using the whole light dome (a), as well as 12 (b) and 6 (c) lights at 51° of elevation. Bottom row: a photograph of the coin illuminated roughly from the same direction as for the renderings (d) and visualizations of the angular errors for 12 (e) and 6 (d) lights.

step via integration [SCS90]. In practical applications, true parallel lighting is hard to achieve [QDW$^+$18]. When using point-like light sources such as LEDs (from here on simply called *point light sources*), both the direction to a given light source and the intensity of incident light depends on the location of the surface point; thus, depth must be estimated together with surface orientation, such that the complexity of the problem is increased [QDW$^+$18].[1]

In literature, a distinction is made between distant-light scenarios (where the parallel light assumption is feasible) and near-light scenarios (where it is not), depending on the size of the imaged scene/object in relation to the distance of light sources. However, this distinction is vague, expressed with adjectives like “near”, “distant”, “large” and “small” without further specification [Cla92, PF14, LND18], or with coarse ratio estimations [XCW15].

The following is written under the premise that when using point light sources, true parallel lighting cannot be achieved in practice and making distinctions between near and distant light sources is misleading. When assuming parallel lighting in PS (whether to reduce complexity in calibration, implementation or computation), a non-zero error must be expected, and the acceptability of this error depends on the requirements of the specific application.

In this section, the relation between light source distance, object size and expected reconstruction error is investigated, as no conclusive account on that topic is found in literature. First, upper bounds for the deviations of light direction and incident intensity for a single light source are determined analytically, as a function of light distance and object size. Combining these bounds with earlier results that relate light source calibration errors to surface normal reconstruction errors, an upper bound for mean reconstruction errors is proposed. The expected errors in surface orientation resulting from different numbers of light sources are then assessed with a Monte Carlo simulation. The results obtained are compatible with the hypothetical upper bounds. Further, an error mitigation strategy is formulated, based on the observation that if depth is known up to a given uncertainty *a priori*, the effective object size and thus the reconstruction error can be reduced to this depth uncertainty. Both the hypothesis on error bounds and the error mitigation strategy are evaluated on real-world datasets. The insights of this work are useful for assessing the feasibility of parallel lighting models in practical applications, or for designing PS image acquisition setups.

### 4.3.1 Object size, light distance and reconstruction error

The ratio of the object size to the distance of light sources is a determining factor for PS reconstruction errors caused by the application of parallel lighting models in a point-light

[1] Of course, even LEDs have an emitting surface with a certain area and are therefore no true point light sources; again, the model error introduced by this inaccuracy will depend on the distance of the light source relative to its size. In this work, however, the effect is considered negligible and not further investigated.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

setup [Cla92, PF14, LND18, XCW15]. In this section, the relationship between object size, light distance and surface normal reconstruction error is investigated in detail.

For the subsequent considerations, global light directions of the parallel lighting model are expressed with respect to a point $\mathbf{c}$, the object center. With $\mathbf{x}$ denoting a surface point subject to reconstruction, the object size $r$ is defined as the maximum distance of any such surface point from the object center.

Given a light position $\mathbf{s}$, the light distance $d$ is defined as the distance from the light source to the object center, $||\mathbf{c} - \mathbf{s}||$. It is reasonable to assume that $0 < r < d$.

Light directions are expressed as unit vectors and with respect to points in object space. Thus, $\vec{s}(\mathbf{c})$ denotes the direction from the object center to the light source (the global light direction in the parallel lighting model), and $\vec{s}(\mathbf{x})$ is the local (correct) light direction from a surface point. Further, $\vec{n}(\mathbf{x})$ is the unit surface normal vector at $\mathbf{x}$.

For ease of discussion, the Lambertian image formation model is adopted:

$$I = max(e\ \rho\ \vec{s} \cdot \vec{n},\ 0) \tag{4.1}$$

where $e$ is the light intensity incident at a surface point (a product of light source intensity and attenuation), $\rho$ is the surface albedo and $I$ is the reflected intensity, here assumed to be equivalent to the image intensity measured. The scalar product of light direction $\vec{s}$ and surface vector $\vec{n}$ yields the cosine of the angle between these vectors.

### Deviations in light direction and intensity

Considering a surface point illuminated by a single point light source, the following parameters are considered constant in the parallel lighting model but generally vary across the surface of an object:

1. incident light direction
2. incident light intensity due to light attenuation, further separable into:
   a) quadratic attenuation due to distance from the light source
   b) radial attenuation due to deviation from the principal direction of anisotropic light sources

First, an upper bound for the angular deviations between globally assumed light directions and the local light directions occurring in a point light setup is determined. Following the definitions at the beginning of this section, all surface points $\mathbf{x}$ are contained in a sphere with center $\mathbf{c}$ and radius $r$. Viewing the schematic in Figure 4.8, it is evident that the maximum angular deviation between light directions $\vec{s}(\mathbf{c})$ and $\vec{s}(\mathbf{x})$ occurs at the tangents to this sphere through $\mathbf{s}$. Expressing the angular deviation at the tangent points via $d$ and $r$ gives the upper bound

$$\delta_s \leq \arcsin \frac{r}{d}. \tag{4.2}$$

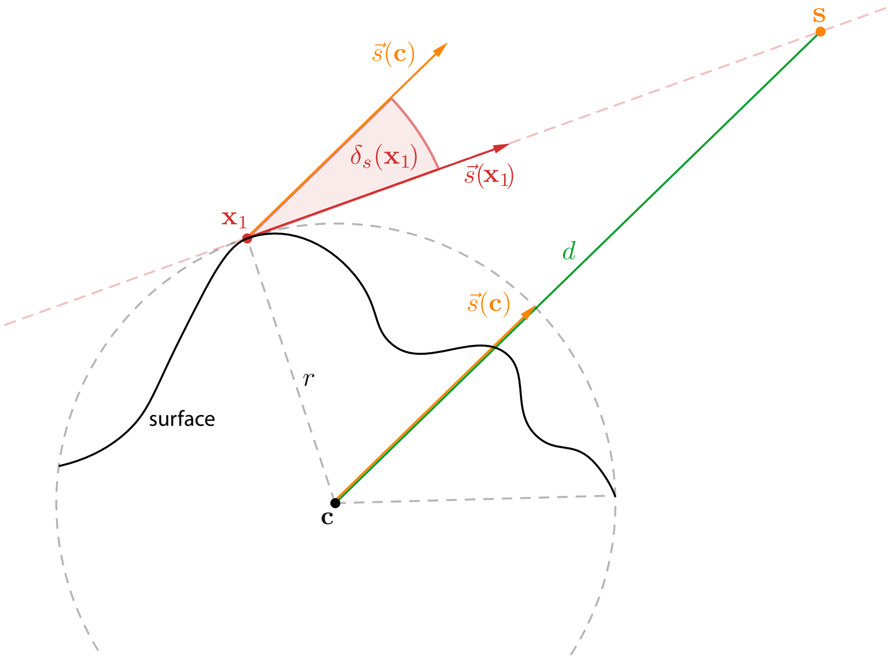


Figure 4.8: Deviations of local light directions when assuming parallel lighting in a point-lighting environment. The theoretical maximum light direction error occurs at the tangent points, here demonstrated with $\mathbf{x}_1$.

In Figure 4.8, this theoretical maximal deviation occurs at point $\mathbf{x}_1$. Note, however, that the existence of such a point is not guaranteed (and not even likely) for a given object and light source. With $d \to \infty$ or $r \to 0$, the upper bound approaches zero, and we arrive at the parallel lighting model.

Next, the variations in light attenuation are discussed. Attenuation is commonly modeled by scaling the light source intensity by a factor

$$a(\mathbf{x}) = \frac{a_r(\mathbf{x})}{a_d(\mathbf{x})} = \frac{\cos^{\mu}(\theta(\mathbf{x}))}{||\mathbf{s} - \mathbf{x}||^2}, \tag{4.3}$$

where the distance attenuation $a_d(\mathbf{x})$ accounts for inverse quadratic attenuation due to distance from the light source. The radial attenuation $a_r(\mathbf{x})$ depends on the angle $\theta(\mathbf{x})$ between light direction $\vec{s}(\mathbf{x})$ and principal direction of the light source $\vec{v}$, and a parameter $\mu$ intrinsic to the light source. For isotropic light sources, $\mu = 0$; otherwise, it is computed as $\mu = -\frac{log(2)}{log(cos(\theta_{1/2}))}$, where $\theta_{1/2} \in (0, \pi/2)$ is the half intensity angle specified by the manufacturer [CRZ+22, QDW+18].

The theoretical object points closest to and farthest from the light source are $\mathbf{c} + \vec{s}(\mathbf{c})r$ and $\mathbf{c} - \vec{s_j}(\mathbf{c})r$, respectively. Thus, distance attenuation is bounded by

$$(d - r)^2 \le a_d(\mathbf{x}) \le (d + r)^2. \tag{4.4}$$

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

The light sources are reasonably assumed to be oriented towards the object center, such that $\theta(\mathbf{c}) = 0$. Thus, the points where the theoretical maximum radial attenuation can occur are, like the points with maximum light direction errors, found at the tangent points defined by $\mathbf{s}$ and the object sphere, and the radial attenuation is bounded by

$$\cos^{\mu}(\arcsin \frac{r}{d}) \leq a_r(\mathbf{x}) \leq 1. \tag{4.5}$$

Therefore, the total attenuation is bounded by

$$\frac{\cos^{\mu}(\arcsin \frac{r}{d})}{(d+r)^2} \leq a(\mathbf{x}) \leq \frac{1}{(d-r)^2}. \tag{4.6}$$

If the light source intensity used in the parallel lighting model corresponds to the incident intensity at object center $\mathbf{c}$, which results from the attenuation

$$a(\mathbf{c}) = \frac{1}{d^2}, \tag{4.7}$$

the relative deviation in light intensity at an object point $\mathbf{x}$ is bounded by

$$\frac{\cos^{\mu}(\arcsin \frac{r}{d})d^2}{(d+r)^2} \leq \frac{a(\mathbf{x})}{a(\mathbf{c})} \leq \frac{d^2}{(d-r)^2}. \tag{4.8}$$

With $d \to \infty$ or $r \to 0$, the bounds tend to 1, and we arrive at the parallel lighting model.

When interested in the greatest possible *relative deviation* from $a(\mathbf{c})$, a choice between the lower and upper bound of Equation 4.8 must be made. As the lower bound is smaller than 1 and the upper bound is greater, the upper bound must be inverted for a meaningful comparison. As it can be shown that

$$\frac{\cos^{\mu}(\arcsin \frac{r}{d})d^2}{(d+r)^2} > \frac{(d-r)^2}{d^2}, \quad 0 < r < d, \tag{4.9}$$

the bound for relative intensity deviations is reduced to

$$\delta_a \leq 1 - \frac{(d-r)^2}{d^2}. \tag{4.10}$$

### Errors in reconstructed surface normals

Assuming the Lambertian reflectance, three input images with linearly independent light directions are necessary for a unique solution [Woo80]. In order to connect error bounds for single light sources (Section 4.3.1) to errors in reconstructed surface normals, previous results on PS error analysis are used. The works of Ray *et al.* [RBK83] and Jiang and Bunke [JB91] relate errors in assumed light directions and measured intensities to reconstruction errors. They independently arrive at the estimation that for a three-light

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

setup and a Lambertian reflector, a 1° error in surface normal reconstruction is caused by a 1° error in light directions or a 1% error in measured intensity [RBK83, JB91]. Their equations indicate that multiple error sources are accumulated additively. These observations are combined with Equations 4.2 and 4.10 to arrive at a hypothesis for an upper bound of the expected normal reconstruction error:

$$\delta_n \leq \arcsin\frac{r}{d} + \left(1 - \frac{(d-r)^2}{d^2}\right)\frac{\pi}{1.8}, \tag{4.11}$$

whereby the factor $\frac{\pi}{1.8}$ simply converts the intensity term from fraction to percentage and from degrees to radians.

In order to assess the validity of this bound for reconstructions from three or more light sources, a Monte Carlo experiment is conducted, in which reconstruction errors resulting from random constellations of surface points, normal vectors and light directions are simulated. The general idea is to randomly sample constellations of surface position, normal vector and light source directions, place virtual light sources at various distances from the object center, compute surface intensities using a point light model, and use these intensities to reconstruct the surface normals under the parallel lighting model; the resulting surface normals are then compared to the true surface normals. A single sample of the simulation is constructed as follows: a random point $\mathbf{x}$ is sampled from the surface of a unit sphere with center $\mathbf{c}$, and equipped with a random normal vector $\vec{n}(\mathbf{x})$ (the object size is thus fixed to 1, and only points at the edges of the object are sampled, where generally the largest deviations are expected, as shown in Section 4.3.1). A random number $m \in [3, 4, ..12, 24, 48]$ of random light directions $\vec{s}_j(\mathbf{c}), j \in [1, 2, ..m]$ is selected[2], with the condition that they do not lead to self-shadowing under the distant light model (*i.e.* $\vec{n}(\mathbf{x}) \cdot \vec{s}_j(\mathbf{c}) > 0$). Virtual light sources $\mathbf{s}_{jk} = \mathbf{c} + \vec{s}_j(\mathbf{c})d_k$ are then placed at a distance $d_k$ from $\mathbf{c}$. Assuming the Lambertian reflectance model, a uniform albedo, equal light source intensities, an anisotropy parameter of $\mu = 1$ (*i.e.*, a Lambertian light source [QDW$^+$18]), and the principal direction of light sources pointing at $\mathbf{c}$, reflected intensities at $\mathbf{x}$ are computed as

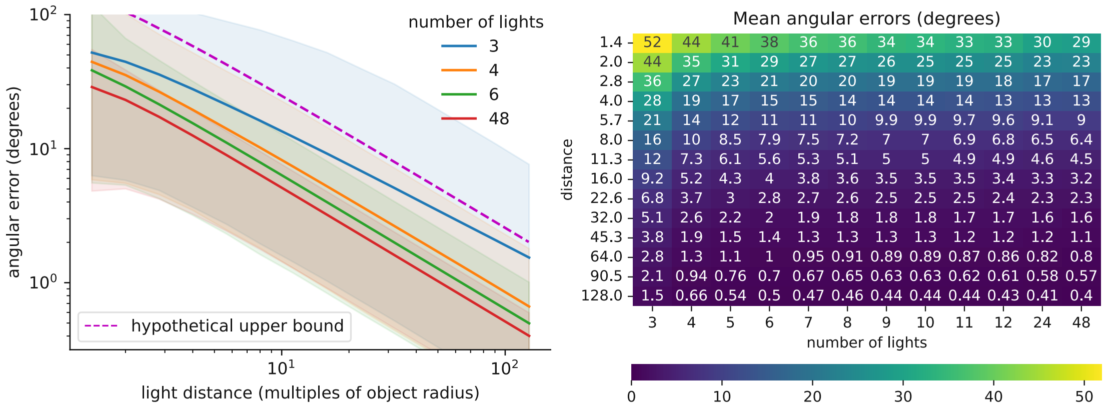
$$I_{jk}(\mathbf{x}) = \vec{n}(\mathbf{x}) \cdot \vec{s}_j(\mathbf{x}) \frac{\vec{s}_j(\mathbf{c}) \cdot \vec{s}_j(\mathbf{x})}{||\mathbf{x} - \mathbf{s}_{jk}||^2}. \tag{4.12}$$

From these intensities, a surface normal is reconstructed using ’vanilla’ PS [Woo80] (assuming parallel lighting) and the angular error with respect to the true surface normal is recorded. This procedure is repeated for 100k constellations of surface positions, normals and light directions, and for $k$ light distances $\sqrt{2}^k, k \in [1, 2, ...14]$ per constellation.[3]

[2] Initial experiments showed that errors start to stagnate at around 12 light sources; thus, the experiments are focused on the range of 3-12 light sources, and the 24 and 48 light configurations are included to illustrate this stagnation.

[3] Expecting a roughly logarithmic behavior of errors, the distance samples are constructed accordingly. The base $\sqrt{2}$ is chosen for more fine-grained results than, *e.g.*, with base 2, but is essentially arbitrary.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a)

(b)

Figure 4.9: Results of the Monte Carlo simulation: Relation between light distance and reconstruction error, split by number of light sources. (a) plots the relationship for a selection of light source counts on a log-log scale, with semi-transparent bands showing the interval between 2.5% and 97.5% percentiles. The hypothetical upper bound is superimposed. (b) lists the mean angular errors for each combination of light distance and light count appearing in the simulation.

The effects of light distance and number of light sources on angular errors of reconstructed normals are shown in Figure 4.9, in the form of mean angular errors over all samples. In Figure 4.9a, the hypothetical upper bound of the errors (Equation 4.11, with $\mu = 1$) is superimposed as a dashed line. The following observations are made:

1. Equation 4.11 gives an upper bound to the mean errors obtained in this simulation. With more than three light sources, also the 95% percentile interval lies below the bound.

2. Errors decrease with an increasing number of light sources and converge eventually.

3. Equation 4.11 explains the mean reconstruction errors up to a factor that depends on the number of light sources (except for the three-light case, which shows a slightly different behavior).

The monotonicity of the results suggests the generalizability for larger light distances and numbers of light sources not evaluated in the simulation.

### Error mitigation via local solutions

The expected errors introduced by the parallel lighting model increase with the object size and decrease with increasing light source distance. In applications where increasing the distance of light sources is not possible, the decrease of effective object size by solving PS piecewise for local surface patches can be considered.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

For such an approach to be applicable, the following conditions must be met:

1. *An approximate object surface is available, with a bounded range of expected deviations.* For example, this approximate surface can be specified by a plane on which an object of interest is placed for imaging, or by a pre-existing coarse 3D model, in applications where PS is used primarily for its ability to capture local detail [HW11, FQW+17, LTBB10, ZPE16, YMH+16].

2. *The incident intensity/attenuation at the approximate surface is known.* This can be achieved analytically if the light source parameters are precisely known, or empirically with the help of a uniformly colored calibration surface. The latter variant is only practical if the approximate object surface is planar (see Section 4.3.2 for a possible implementation).

Note that these conditions are plausible for graphical heritage applications (see [YMH+16, LTBB10] and the examples given in Section 1.1.1).

For each patch, a local object center $\mathbf{c}_p$ is defined on the approximate surface, and the local light directions and incident intensities with respect to $\mathbf{c}_p$ are determined. As illustrated in Figure 4.10, the effective object size $r_p$ (and thus the expected reconstruction error) results from the horizontal expansion of the patch chosen and the resultant depth range covered by the patch. The greater the horizontal expansion of the entire object imaged relative to the expected depth range, the greater the potential reduction in effective object size.

### 4.3.2 Experiments

In Section 4.3.1, the reconstruction errors caused by a parallel lighting model in a point light setup are analyzed and based on the findings, an error mitigation strategy via local solutions is justified. These theoretical results are now experimentally tested, using the *Diffuse objects* dataset (see Section 4.1.2). For efficiency, only six light sources arranged in a circle at an elevation angle of about 50° w.r.t. the dome center are used for these experiments; as determined in Section 4.2, the impact of this reduction on reconstruction quality is limited.

#### Surface normal reconstruction

As in this setup, both light distance and object sizes are fixed, the strategy of piecsewise solutions formulated in Section 4.3.1 is employed to vary the effective object size and observe the effect on surface normal reconstruction error. Here, the planar base surface on which the test objects are placed serves as the required approximate surface; the expected depth range is given by the depth covered by the entire test object, measured from the ground truth 3D models (see Section 4.1.2). The approach also requires the incident light intensities on points of the base surface. As only the positions but not the emission characteristics (principal direction and anisotropy) of the light sources are

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

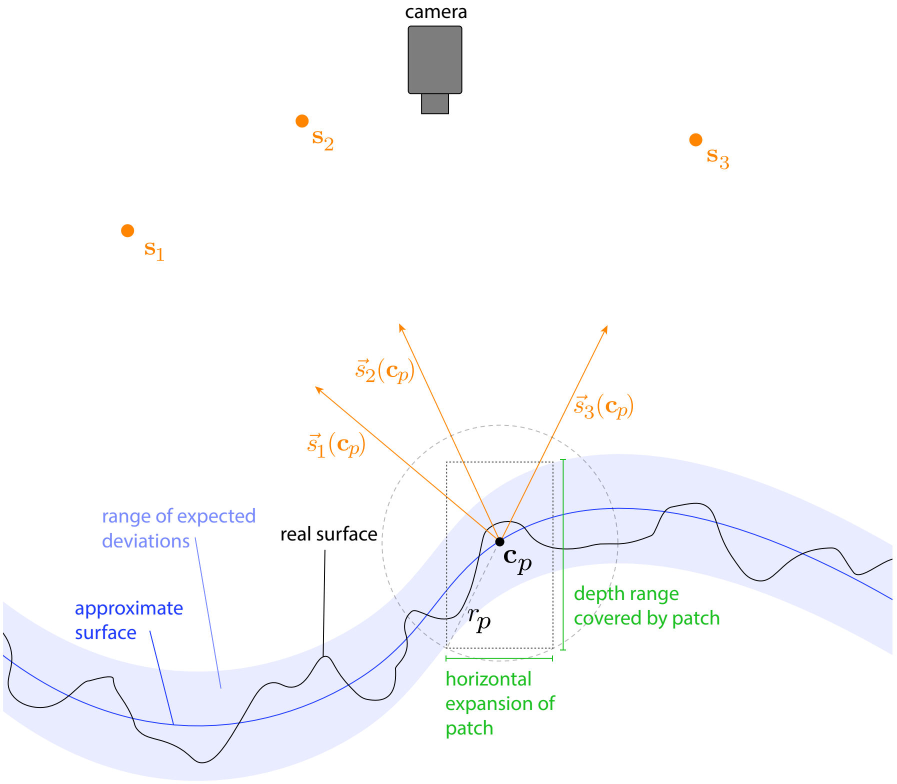

camera
$\mathbf{s}_2$
$\mathbf{s}_3$
$\mathbf{s}_1$
$\vec{s}_2(\mathbf{c}_p)$
$\vec{s}_3(\mathbf{c}_p)$
$\vec{s}_1(\mathbf{c}_p)$
range of expected deviations
real surface
$\mathbf{c}_p$
approximate surface
$r_p$
depth range covered by patch
horizontal expansion of patch

Figure 4.10: Patchwise PS. With an approximately known base surface (blue line), PS can be solved locally for small patches. For the calculation of expected errors, the potential depth values covered by the patch must be considered. For ease of discussion but without loss of generality, an orthogonal projection is assumed in this illustration.

known, incident intensities are calibrated using images of the reference plane. Using the reflectance model in Equation 4.1, the incident intensity $e(\mathbf{x})$ at a given point $\mathbf{x}$ on the base surface is expressed as

$$e(\mathbf{x}) = \frac{I_j(\mathbf{x})}{\rho \; \vec{s}_j(\mathbf{x}) \cdot \vec{n}(\mathbf{x})}. \tag{4.13}$$

The incident intensity combines the light source intensity and attenuation, and is computed using the intensity $I_j(\mathbf{x})$ measured in the image of the surface with the $j$-th light source active, the surface albedo $\rho$, which is constant (and can be set to an arbitrary non-zero value, *e.g.*, 1), and the local light direction $\vec{s}_j(\mathbf{x})$ and surface normal $\vec{n}(\mathbf{x})$, which are straightforwardly computed knowing the locations of reference plane and light sources.

Surface normals are computed for different subdivision levels of the input images. On the first level, the whole input image is processed; then, the image is successively divided into square patches with a side length $\hat{l}_p$ which is halved in each iteration. For each patch, the object center $\mathbf{c}_p$ is set at the horizontal center of the patch and at the depth

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

of the base plane. The effective object size is approximated as

$$r_p = \sqrt{\frac{{l_p}^2}{2} + \Delta z^2}, \tag{4.14}$$

where $\Delta z$ is the depth range covered by the whole test object (constant) and $l_p$ is the image patch size $\hat{l}_p$ projected onto the base plane. An exception is made for the first level, where the $l_p$ is determined by the dimensions of the test object (which in most cases does not cover the whole imaged area).

For each patch, pixel-wise surface normals are computed with 'vanilla' PS [Woo80] (*i.e.*, assuming parallel lighting and Lambertian reflectance, and not accounting for shadows), whereby light direction and incident intensity are determined w.r.t. $\mathbf{c}_p$ and used equally for all pixels of the patch.

### Results

The surface normals gained in the experiments are evaluated by means of angular errors w.r.t. the ground truth normals. Only the areas within the object masks are considered for evaluation. The mean errors, grouped by test objects and relative object size, are shown in Figure 4.11; for reference, the hypothetical upper bound and the simulation results for six light sources (see Section 4.3.1) are superimposed. Note that while the experiments are conducted with equal patch sizes in image space for all test objects, the resultant effective object sizes differ, as each object covers a different depth range. In Figure 4.12, the different effects of patch size on reconstruction error can be observed qualitatively. As non-overlapping patches are chosen for these experiments, discontinuities are visible at their borders - smooth normal maps can be obtained by choosing overlapping patches and blending the patch-wise results. The high error values visible in some areas of the *monkeys* object that stay constant over all object sizes are related to cast shadows, which are not handled in the reconstruction method employed.

On average, the errors are higher than in the simulation, but still below the hypothetical upper bound; only at large light distances ($\gtrapprox 100\ r$), the errors appear to converge towards a residual error, instead of zero (seen best with the *flatfield* object, where $\Delta z = 0$ and thus the effective object size could be reduced the most). This is expected behavior, as the lighting model is not the only source of error in real-world PS. It is worth noting the light sources are evenly distributed and at a reasonable elevation angle (see [SC06] and Section 4.2), while in the Monte Carlo simulation, the light directions are chosen randomly. The results are therefore not representative of all possible light arrangements, but only of a well-planned acquisition setup. In the experiments described above, lens distortion was not corrected, as due to the low distortion of the camera system used (see Section 3.1) no severe effects are expected. Generally, however, undistortion of input images is recommended as a pre-processing step in order to obtain correct estimates of local light directions $\vec{s}_j(\mathbf{x})$.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

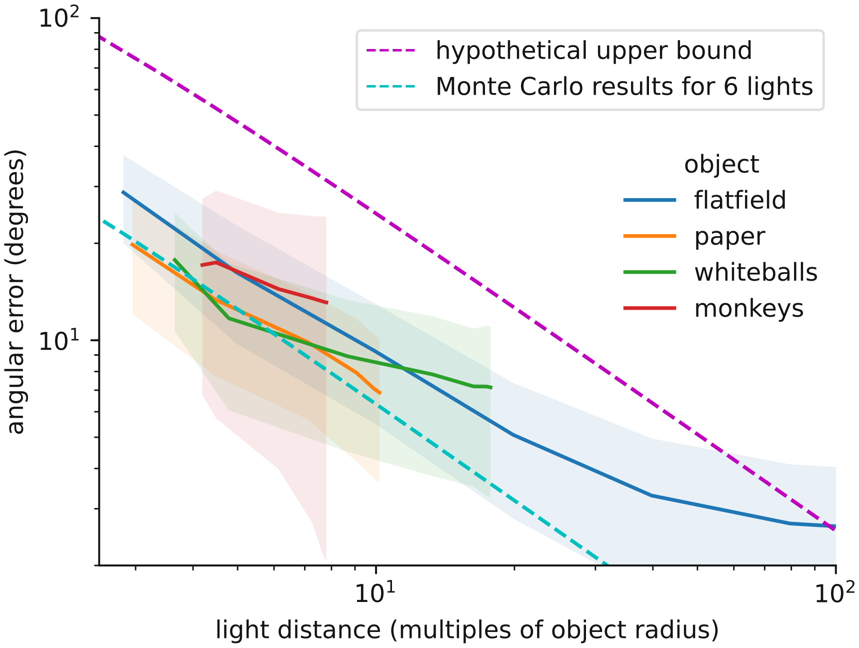


Figure 4.11: Mean angular errors obtained in the lab results, split by test object. Standard deviations are shown as transparent bands. The hypothetical upper bound (Equation 4.11) and results from the Monte Carlo simulation are shown as dashed lines.

### 4.3.3 Discussion

In this section, the errors resulting from the usage of a parallel lighting model in a point light PS setup are analyzed as a function of object size and light source distance. After deriving upper bounds for errors in local light direction and intensity, a hypothesis for an upper bound of normal reconstruction errors is formulated, based on previous error analysis results [RBK83, JB91]. This hypothetical upper bound is shown to hold for mean angular errors obtained from both a Monte Carlo simulation and for the errors obtained in laboratory experiments. In these experiments, the applicability of the proposed error mitigation strategy based on local solutions is also demonstrated.

From these results, implications for the feasibility of parallel lighting models in practical applications can be deduced. When planning PS image acquisition, Equation 4.11 gives an estimate for the surface normal reconstruction errors that should be expected for a given object size and light distance - the acceptability of this error then depends on the requirements of the application. Note that while this estimate is pessimistic when considering ideal imaging conditions, it only includes errors introduced by using a parallel lighting model in a point light setup - other sources of errors (*e.g.*, sensor noise, cast shadows, inaccuracies of the reflectance model) must be considered additionally.

The error mitigation strategy presented is beneficial for use cases in which the uncertainty

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

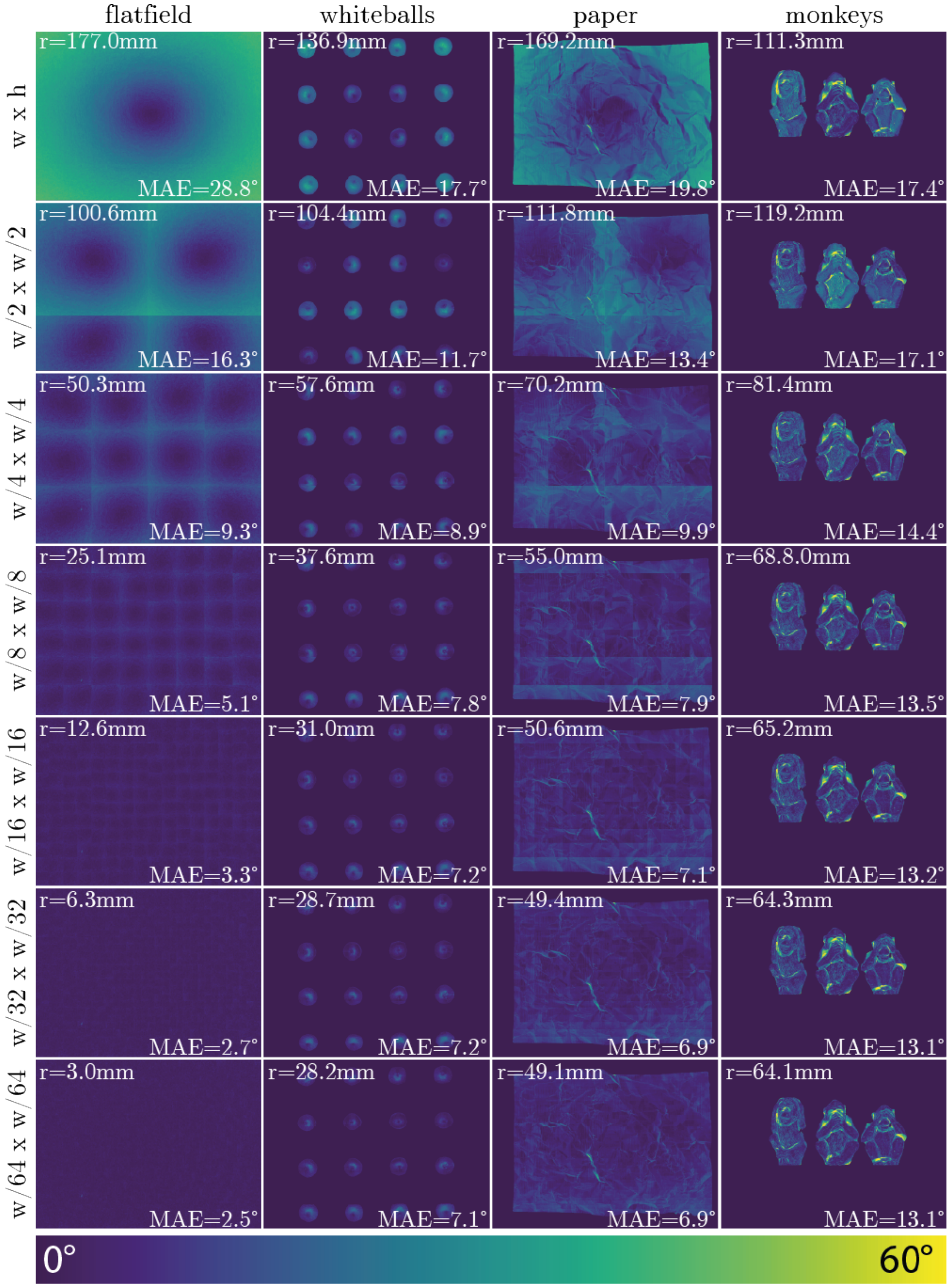


Figure 4.12: Angular error maps obtained in the lab results. Columns correspond to different test objects, rows correspond to different patch sizes, decreasing from top to bottom. Effective object sizes and MAEs are inscribed in each result image.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

in depth is much smaller than the horizontal extent of the object imaged. This occurs in numerous applications where PS is used for its merit in recording local surface details, for example: in (historical) document analysis, PS is used for capturing dry-point ruling lines or tracing lines [VHW$^+$18], fiber structures of substrates [KE15] or impressions left by a pen [MC03]; in industrial surface inspection, PS is used for defect detection [SBAG22, WZSE15, KJW13, FSSM05] or surface characterization [JYP07, ZPE16]. In contrast to the approaches discussed in Section 2.2.2, no additional computational complexity is introduced with respect to original PS [Woo80].

In conclusion, this work contributes to the understanding of errors in PS and serves as a reference to assess the feasibility of classical PS in a point light setup. Questions left unanswered in this work include the applicability of the results to different reflectance models and the reasons for the aberrant behavior of three-light constellations in the Monte Carlo simulation.

## 4.4 A case study: Recovering poetry by W. H. Auden

After the abstract discussions in the preceding sections, the reader is now treated to a demonstrative example of interdisciplinary work, in which previously described results are applied and PS actually contributes to investigations on graphical heritage [BFM23].

### 4.4.1 Background

W. H. Auden (1907-1973) was a highly influential British-American poet in the twentieth century (*e.g.*, 1948 Pulitzer Prize for Poetry for “The Age of Anxiety”, 1956 National Book Award for “The Shield of Achilles”, professorship of Poetry at the University of Oxford from 1956 to 1961). From 1958, Auden owned a house in the Austrian village of Kirchstetten, where until his death in 1973 he spent up to six months each year and wrote most of his later poetry [Qui15, Qui13].

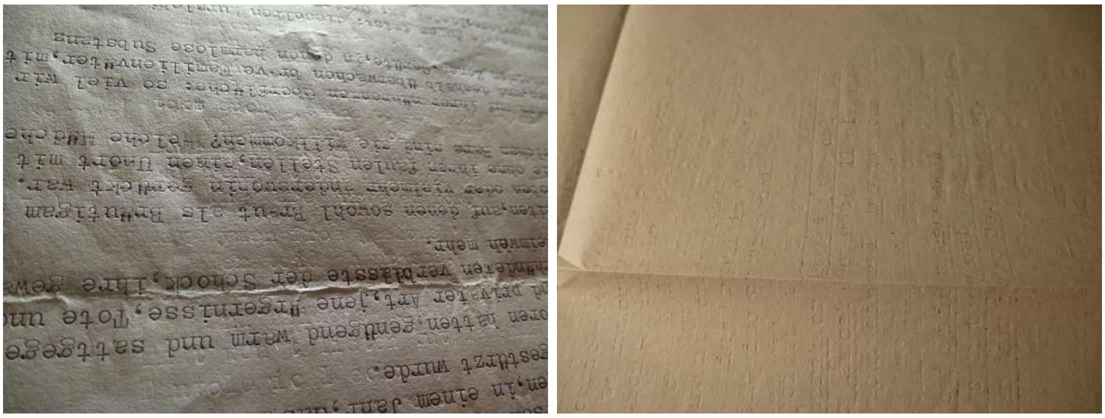
The *Auden Musulin Papers* project[4] sheds new light on Auden’s life in Austria by investigating a previously inaccessible, private collection of his ’working correspondence’ with Welsh-Austrian writer Stella Musulin [Mus95], along with literary papers, drafts of speeches, and photographs.

In the collection, two documents containing inkless typewriter impressions were found. In times when the typewriter was a key writing technology in literary production [Lyo21], it was standard practice to insert a ‘backing sheet’ into the typewriter below that sheet onto which was written, in order to alleviate the type’s impact on paper and rubber platen [Bai90]. Thereby, imprints of the typed text may be left on the backing sheets; and Auden has reused such sheets for his correspondence and literary writing [BFM23].

[4] The project is based at the Austrian Centre for Digital Humanities and Cultural Heritage of the Austrian Academy of Sciences and funded by the Austrian Science Fund (FWF), grant number P 33754.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b)

Figure 4.13: Raking light images of (a) the typescript of 'Joseph Weinheber' from 28 April 1965 and (b) a letter to Stella Musulin from 10 June 1969. Photographs: Timo Frühwirth.

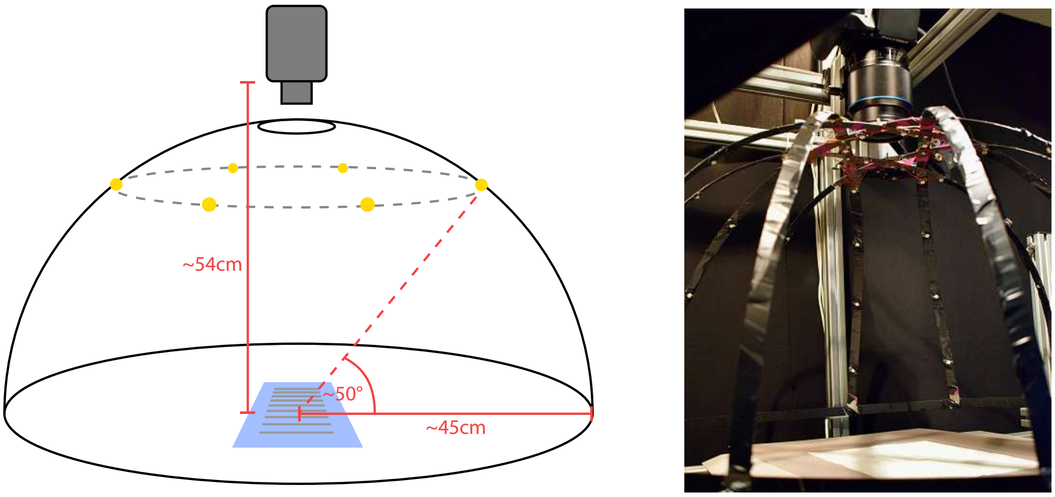
Initial attempts to visualize the indented letters by means of raking light rendered the texts visible, but largely undecipherable. In one of the documents (Figure 4.13a), the impressions are overwritten by the inked (as well as impressed) lines from another typescript of the poem 'Joseph Weinheber'. The sheet used for the second document, a brief handwritten letter to Stella Musulin, was used as a backing sheet twice, in different orientations - thus, the page contains overlapping lines of indented text, one of them in mirror-image (Figure 4.13b). In both cases, the impressed texts cannot be fully retrieved from the multiple, multidirectional layers of writing and overwriting by means of raking light only.

### 4.4.2 Photometric stereo reconstruction

Images were acquired with the system described in Section 3.1; following the insights from Section 4.2, only a subset of six LEDs in a circular arrangement at approximately 50° elevation angle from the center of the hemisphere was used (see Figure 4.14b). With an imaging distance of ca. 540 mm, a surface resolution of approximately 41 pixels per millimeter (~1030dpi) was measured - with this configuration, only half of the original A4 format fits in a single image.

Light source positions were calibrated with the approach described in Section 3.3. Assuming an object size $r \approx 128$ mm (half the diagonal of an A5 page) and mean light distance from the object center $d \approx 450$ mm, Equation 4.11 gives an upper bound of $\delta_n \leq 65.5°$ for surface normal errors introduced by the parallel lighting model. However, due to the approximately planar base shape of the leaves, the error mitigation approach described in Section 4.3.1 can be applied very effectively: solving PS for local square patches with $r_p = d/100$ reduces the expected errors to 2.6°. Measurement of incident intensity at patch centers is effectively done with a piece of flat gray cardboard, like demonstrated in

TU Bibliothek WIEN Your knowledge hub



(a) (b)

Figure 4.14: Acquisition setup: (a) The schematic illustration shows the locations of light sources (yellow), camera (grey), and document (blue). (b) The actual setup for imaging the letters (ambient light is eliminated during acquisition).

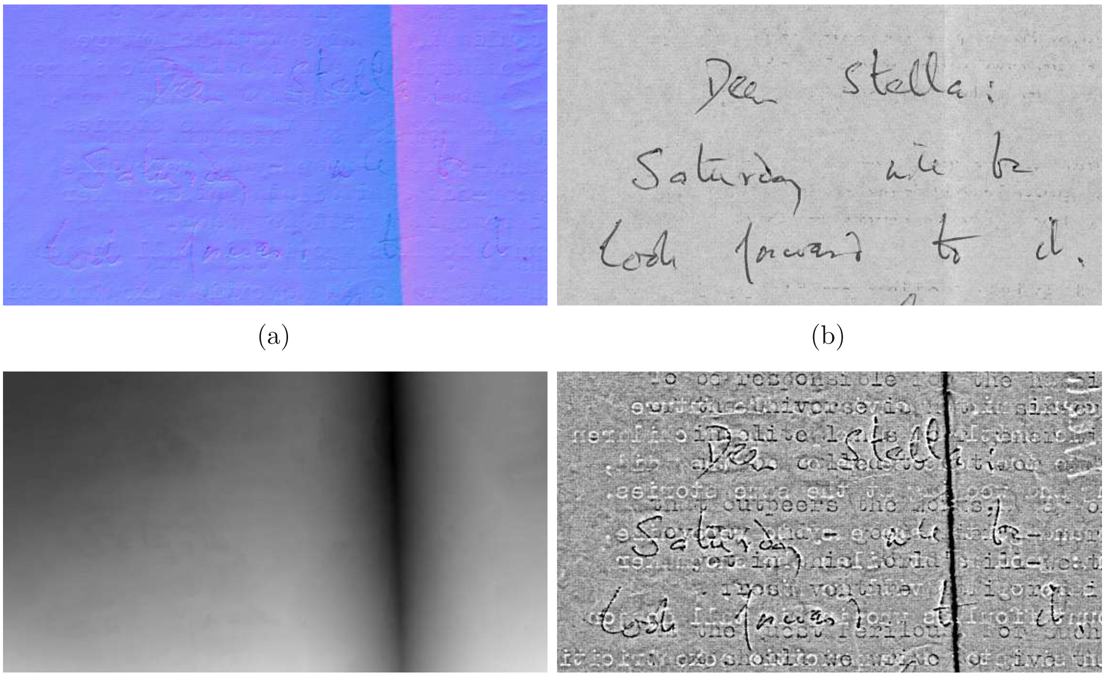
Section 4.3.2 - when placed on the same surface, not only the shape, but also the diffuse reflectance is similar to the imaged sheets of paper.

In order to achieve a seamless combination of the partial results, neighboring patches are overlapped by 25% and blended via feathering [GK16]. Surface depth is obtained via integration of the blended normal maps, using the method of Simchony *et al.* [SCS90]. An example of a resulting normal map, albedo map, and depth map of a page section is shown in Figures 4.15a-4.15c.

Evidently, none of these naive visualizations are particularly useful for reading the indented text lines. Especially in the depth map, variations introduced by typewriter impressions are very small compared to global depth variations of the paper (folds, bends, *etc.*). Thus, these large-scale variations are removed by means of a high-pass filter, implemented as a subtraction of a Gauss-filtered version of the depth map (with $\sigma_{hp} = 0.25mm$, determined experimentally) from the original. For visualization, the resulting depth values are mapped to grey values so that $\mu - 3\sigma$ corresponds to the minimum intensity value of an image and $\mu + 3\sigma$ corresponds to the maximum (where $\mu$ is the mean depth and $\sigma$ is the standard deviation). An example of the resulting image is shown in Figure 4.15d, in which letters darker than the background correspond to impressions from the viewer's side (recto) and letters brighter than the background correspond to impressions from the opposite side (verso). Note that the handwriting and one set of typewriter impressions are shown in similar grey values. To achieve a clear distinction between the originally colored text sources and the inkless impressions, the depth map can be combined with the albedo map. Examples of such combined visualizations are shown in Figures 4.16a (same page section as in Figure 4.15) and 4.16b

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



(a) (b)

(c) (d)

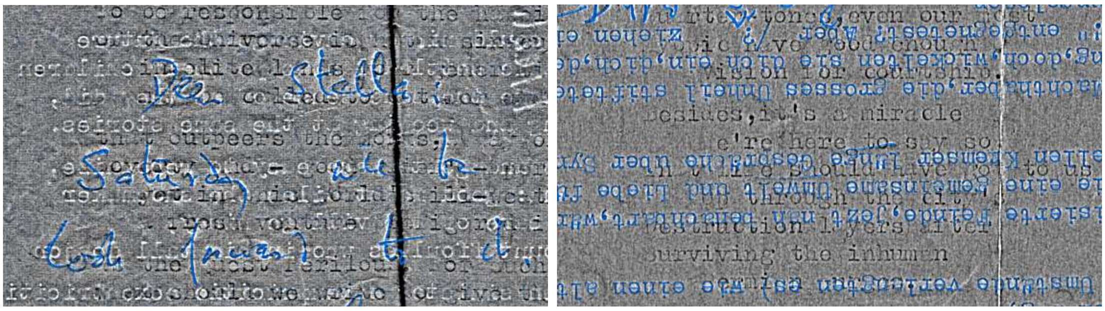
Figure 4.15: Visualizations of PS results: (a) Normal map, with x-, y-, and z-components of the surface normal vector at each pixel position being encoded in the red, green, and blue channels of an RGB image; (b) albedo map, *i.e.*, surface reflectance as a material property; (c) depth map, encoding the distance from the camera to the imaged surface, with darker colors corresponding to greater distances from the camera; (d) high-pass filtered depth map.

(a section of the typescript of 'Joseph Weinheber'); the full visualizations of the two sheets imaged from their front and back sides are available on **zenodo** [Bre23].

### 4.4.3 New insights into Auden's writing practice

The two scenarios anticipated as outcomes of reconstructing the indented texts are (1) that the retrieved typescript matches a document now housed in an archival collection, or (2) that the reconstruction reveals a previously unknown typescript. Both scenarios are exemplified by the two documents of this case study: while the impressions in the letter to Musulin from 1965 correspond to an unpublished version of the poet's 'Epistle to a Godson' that corresponds to a document now held by the Bodleian Libraries, the 'Joseph Weinheber' typescript contains an entirely unknown early version of the poem 'Epithalamium'.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b)

Figure 4.16: Final visualizations in the form of high-pass filtered depth maps combined with albedo maps, where inked parts of the surface appear in blue: (a) letter to Stella Musulin, (b) German typescript of 'Joseph Weinheber'.

### An Unpublished Version of Auden's Poem 'Epistle to a Godson'

W. H. Auden's poem 'Epistle to a Godson' was first published in June 1969 in *The New York Review of Books* [Aud69], and appeared later in his 1972 book of poems *Epistle to a Godson and Other Poems* [Aud72]. Based on the shape of the text as well as on typewritten deletions and additions, the impressions found in the letter have been identified as originating from a typescript of the poem sent by Auden to E. R. Dodds, an Oxford scholar and friend of the poet. The letter containing the poem was sent from Kirchstetten on 10 June 1969, but the poem itself is undated. Combining textual features – verbal variants, spelling, and punctuation – of both the 1969 and 1972 texts, it can be described as a preliminary, intermediate version between the two, thus evidencing a swift mode of poetic revision.

As an impression of the poem was found on a letter, of which the sending date is known through postmarks on the containing envelope, it is now known that both original typescript and impressions were sent on the same day (10 June 1969). This additional piece of information suggests that both the typescript of 'Epistle to a Godson' and the letter to Stelle Musulin were written shortly before this date.

### A Lost Early Version of 'Epithalamium'

On 28 April 1965, Auden had sent to Musulin the typescript of a German-language prose translation of his poem 'Joseph Weinheber'. It was typed on a backing sheet with impressions of an early version of his poem 'Epithalamium', written for the occasion of his niece Rita Auden's wedding. 'Epithalamium' was first published in *The New Yorker* on 31 July 1965 [Aud65]. The restored poem substantially deviates from that text and cannot be related to any published versions of the poem or any poetic typescript held by any public archive - it has survived only in the form of indentations.

The discovery of the indented poem has also yielded contextual information related to Auden's poetic practices of composition and revision. Auden submitted the New Yorker poem for publication on 29 May 1965 [Men22]. The indented version, however, must have been written before 28 April, when the respective backing sheet was sent in a letter. A further detail, however, has led to the assumption that this version was typed exactly on 28 April 1965: On that day, Auden wrote in his pocket diary: 'Epithalamium done but for names' [Men22]. The restored poem includes 49 lines as opposed to the 63 lines of the poem published in The New Yorker; It breaks off immediately before an enumeration of family names of bride and bridegroom. It would seem plausible that Auden made inquiries about those names before noting in his diary, on 5 May, that he had 'received names' [Men22]. This suggests that, after immediately reusing the backing sheet, he may have kept this incomplete version until he would be able to finish the poem (and perhaps disposed of it afterwards).

For a more detailed account of backgrounds and new insights, the reader is referred to Brenner *et al.* [BFM23].

### 4.4.4 Implications

Beyond the specific results outlined above, the case study has wider implications with regard to the use of PS in twentieth-century literary studies. The use of backing sheets was standard practice during the typewriter era, and, through establishing spatio-temporal relationships between the texts contained in the impressions of these sheets and the messages overwriting them, the textual information can be enriched by con-textual information.

More generally, the case study contributes to a re-conceptualization of sheets of paper as three-dimensional objects in the research of documents from the typewriter age. The three-dimensionality of older near-flat text and image carriers (*e.g.*, medieval parchment manuscripts, renaissance drawing paper, eighteenth-century copperplate prints, Sanskrit palm-leaf manuscripts, or papyri) is more widely acknowledged and researched with PS and related techniques [VHW$^+$18, End19, Bar22, KE15, Piq17]. This case study, however, has pioneered PS in the context of twentieth-century literature.

## 4.5 Summary

For PS reconstruction of quasi-flat surfaces, circular light arrangements with an elevation angle between 32° and 51° result in reconstruction errors comparable to dome-shaped arrangements, using a fraction of light sources. The expected errors introduced by the assumption of parallel lighting in point light environments are expressed in terms of light source distance and object size (Equation 4.11). When solving PS piecewise with local light directions under knowledge of an approximate surface, the effective object size is reduced and thus errors are mitigated. These insights are used when reconstructing poetry by W. H. Auden from inkless typewriter impressions - visualizations based on depth maps allow a near complete reading of the respective texts and their identification as unpublished intermediate versions of known poems.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

CHAPTER 5

# Evaluating graphical heritage imagery

How many roads must a man walk
down before you call him a man?

*Bob Dylan*

In the preceding sections, acquisition and low-level image processing methods for multi-light imaging were discussed. Their immediate goals - *e.g.*, producing sharp images or correct surface normals - are evaluated straight-forwardly. However, returning to the high-level aim of this work, *i.e.*, "recording and visualizing surface variations of physical objects, such that the results serve as evidence for the discovery and characterization of graphical heritage" (see Section 1.1.2), it becomes evident that evaluating any image generating method with respect to these goals is not trivial - this chapter is dedicated to this problem.

To limit the scope of investigations, the particular problem of assessing imagery with respect to its use for deciphering textual contents (*i.e.*, its *legibility*) is treated. Although the restoration of degraded materials via imaging and image restoration has received much interest in scientific literature [ACD18, ECK11, GCM$^+$17, Min18, PDG$^+$17], a generally applicable objective metric for evaluating the results (for the property of human legibility) is missing [Hol21, p.18]. Consequently, the evaluation of proposed text restoration approaches is commonly based on expert ratings, the demonstration on selected examples or case studies [ECK11, Min18, PDG$^+$17]. This practice is unfavorable for the research field: it does not allow for an automated evaluation on large public datasets, such that an objective comparison of different approaches, and thus measurable progress, is impeded. Another use case where a quantitative legibility metric would be beneficial is automated parameter tuning for producing an optimally readable version of a specific document [SNAMC18, SF11, GC16], where the meaning of 'parameter' can range from the choice of a processing algorithm up to fine-tuning of method-specific constants.

For a meaningful application in the area of degraded text restoration, I propose the following requirements that must be fulfilled by a quality metric:

**R1 Rank correlation to human legibility.** The transcription of heavily degraded manuscripts is performed by experienced scholars and currently inconceivable using automated methods. It is thus the usefulness of an image for a human reader that ultimately defines its quality; machine readability and aesthetical properties are subordinate. A quality metric should thus be able to order images with respect to legibility the same way a human reader would.

**R2 Robustness to script and language.** Considering the variety of scripts and languages found in ancient documents, it would be desirable for a quality metric to be largely agnostic of them. This is especially relevant for rare scripts of which not many samples are available.

**R3 Independence from reference.** In order to make a judgment, a quality metric should require no information other than the image being judged (*e.g.*, labeled pixels or prior knowledge about the text contained). This extends the applicability of the metric from dedicated datasets to arbitrary images of unknown documents.

**R4 Applicability to non-binarized images.** Binarization is a typical pre-processing step for Optical Character Recognition (OCR), but not desirable for human observers. For historical handwritten documents, the risk of information loss due to binarization artifacts is too high [YSB+15], especially when the text is heavily degraded.

As a first approximation to the problem, an approach introduced by Giacometti *et al.* [GCM+17] based on a dataset of artificially degraded manuscripts is further investigated. After concluding that this approach inherently violates Requirement **R3** and its conformity to Requirement **R1** remains to be shown, an own dataset of Subjective Assessments of Legibility in Ancient Manuscript Images (SALAMI) is introduced. Based on this, a framework and baseline for the evaluation of quantitative image quality estimators is proposed.

## 5.1 Artificially Degraded Manuscripts for Legibility Assessment

The problem of quantitatively evaluating text legibility in images can be considered a special case of Image Quality Assessment (IQA). For degraded graphical heritage imagery,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

a reference image with intact contents is typically not available, such that full-reference IQA (generally a more well-posed problem [LW18]) is not applicable.

However, Giacometti *et al.* propose a way to perform quality assessment for graphical heritage imagery in a full-reference setting by means of a dedicated dataset [GCM+17]. They cut patches from an 18th-century document written with iron gall ink on parchment and acquired Multispectral (MS) images before and after artificial degradation by various treatments, *e.g.*, scraping, staining with various substances, or exposure to radiation. The resulting dataset [GCM+15] consists of 23 manuscript patches, of which 20 were subject to a different treatment each and three were left untreated as control images. Two of the patches were imaged from both sides (because both sides contain text), giving a total of 25 samples.

The dataset is then used to conduct a study on the performance of MS imaging and post-processing techniques for recovering information lost in the degradation process. The result images are compared with the untreated originals. The authors employ mutual information [VWI97] as a similarity metric.

This work is significant because, to the best of my knowledge, it resulted in the first dataset systematically documenting the effects of degradation processes on the spectral response of written text and potentially enabling an objective evaluation of attempts to restore the original information. However, it has several restrictions for a broader application: First, the number of samples is small and, as all samples are taken from the same manuscript, there is no variation in substrate and ink composition. Second, the accompanying publication [GCM+17] fails to conclusively show that comparison with the original image is a valid method to assess the quality of text restoration; although plausible results are shown for selected examples, the generality of the results is not discussed. However, this would be a prerequisite to legitimate further studies of this kind with a higher number of samples and greater variation. In the following, the results described in the original paper are reproduced and extended in order to further investigate this latter issue.

### 5.1.1 Pre-processing

The published dataset [GCM+15] contains MS images acquired with a monochromatic scientific camera as well as color images. In the following, only the monochromatic images are considered. For each sample, 21 spectral layers from 400 nm to 950 nm are available for the untreated and treated variants. The layers are intensity normalized [MGC+13] and inter-registered; however, the treated images are not registered to the untreated ones. Also, a set of results from dimensionality reduction methods is provided for each sample; they are registered to the untreated variants, but far from pixel-accurately, which biases quantitative comparisons. Giacometti *et al.* [GCM+17] do not describe against which particular images the processed versions are compared. For these reasons, the dataset is pre-processed before using it for continuing experiments:

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

1. From the untreated image, a panchromatic image is created by averaging the layers in the visible range ($400nm < \lambda < 700nm$). For the sake of simplicity and uniformity, these panchromatic images will serve as references for registration and comparison, and will from here on be referred to as *reference*.

2. One layer of the treated sample is registered to the reference using a deformable registration framework for medical image processing [KSM+10, SBL+14]. The 800 nm layer was chosen for that purpose, as a visual assessment showed that it shares most of the textual information with the untreated images for the majority of degradation types. A deformable registration approach is necessary due to deformations of the parchment resulting from the treatments. The remaining treated images are registered using the transformation found for the 800nm layer.

3. Panchromatic images and registered treated images are cropped to 900 × 900 pixels.

4. To produce test images that can be compared with the reference, the cropped registered treated images are processed with five common (but arbitrarily chosen) dimensionality reduction methods: principal component analysis, independent component analysis, factor analysis, truncated singular value decomposition and k-means clustering. From each method, five components are extracted, leading to a total of 25 processed variants for each sample, from here on referred to as *processed images*.

The three samples treated with heat, mold and sodium hypochlorite could not be registered satisfactorily due to their condition and were thus omitted, leaving 22 samples for investigation. The resulting modified version of the dataset is available online [Bre19].

### 5.1.2 Comparison metrics

The images retrieved from dimensionality reduction methods visualize statistical dependencies rather than measured intensity values, such that contrast, mean brightness and *polarity* (in our case referring to dark text on light background versus light text on dark background) of these images typically deviate from the original photographs [GCM+17]. Therefore, any comparison metrics that rely on absolute intensity differences, such as the Mean Squared Error or Peak Signal To Noise Ratio, are unsuitable for this application. Instead, metrics that provide a measure of structural similarity and are insensitive to contrast and polarity are required.

Viewing the pixel positions as observations and the intensity values of the compared images as observed variables, statistical measures of dependence such as the Pearson Correlation Coefficient (PCC) and Mutual Information (MI) between the variables (*i.e.*, images) are available as relevant comparison metrics. While MI, which Giacometti *et al.* employed in their work [GCM+17], can be used as-is, reversed polarities result in negative PCC values such that the absolute value is used as a score.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Alternatively, established no-reference IQA metrics emphasizing structural similarity like the Structural SIMilarity index (SSIM) [WBS+04] and Visual Information Fidelity (VIF) [SB06] are available. Although these metrics are not agnostic of contrast, its influence can be adjusted with a parameter for SSIM, while VIF actually rewards images with higher contrast than the reference. To make the methods invariant to polarity, $max(\varphi(I_{ref}, I_{test}), \varphi(I_{ref}, \neg I_{test}))$ is used as a comparison score, where $\varphi$ denotes either SSIM of VIF between two images and $\neg$ is the image complement.

More advanced full-reference IQA metrics (*e.g.*, based on learning) are consciously omitted for these initial experiments as they would introduce unnecessary complexity.

### 5.1.3 Experiments

In order to reproduce previous results [GCM+17] and investigate the feasibility of comparison with an intact original as a measure for legibility, each processed image is compared with the reference using MI as well as the adapted variants of PCC, SSIM and VIF described above. The use of additional similarity metrics allows to observe if the choice of metric significantly influences the results. The scores are then used to create rankings of the processed images for each sample, allowing to visually assess their plausibility.

In addition, the influence of contrast enhancement on the respective scores is evaluated experimentally: For each sample, the first five principal components (showing varying degrees of initial contrast) were subjected to Contrast Limited Adaptive Histogram Equalization (CLAHE) with varying clip limits to monitor the influence on the different scores.

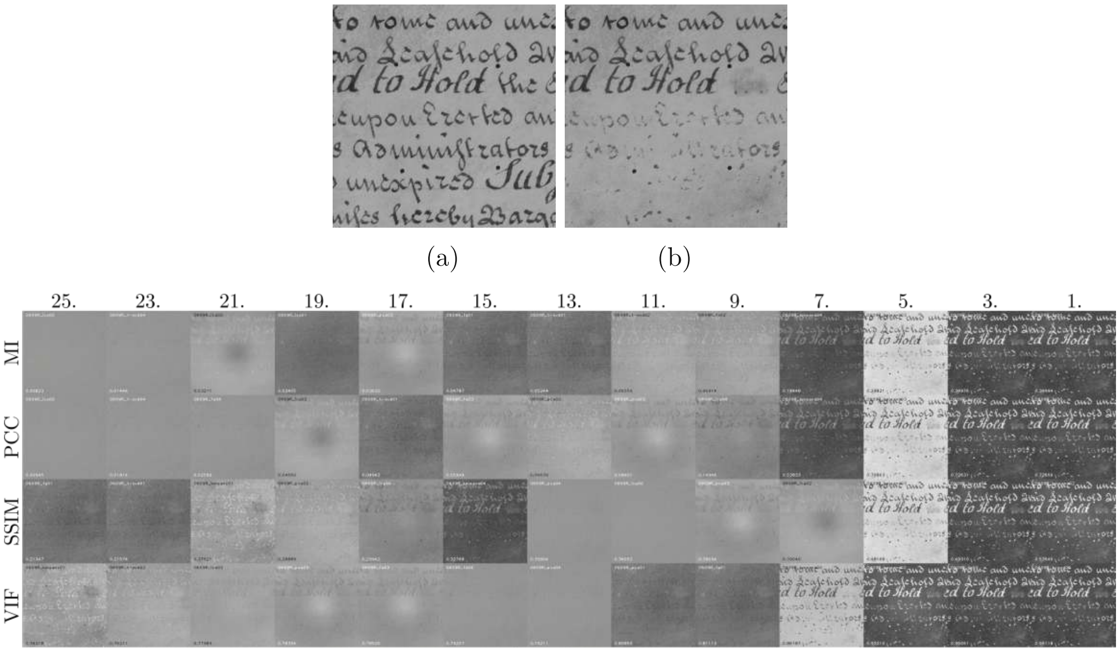
The full results of the experiments as well as relevant source code can be accessed online along with the pre-processed version of the dataset [Bre19].

### 5.1.4 Results

Visually assessing the processed image variants ranked by the employed comparison metrics generally confirms the assumption that similarity to a non-degraded reference image correlates well with the legibility of text. The example shown in Figure 5.1 is representative of the remaining samples, where similar situations are observed.

The rankings derived from different similarity metrics are well correlated, with MI and PCC showing the strongest agreement. This is comprehensible when visually assessing rankings like in Figure 5.1c, and also manifests in the correlation matrix of the different metrics, which is shown in Table 5.1.

One could assume that the good scores of the highest-ranked images are due to their high contrast; this general assumption, however, is readily disproved. Experiments with different levels of generic contrast enhancement show that it has no positive effect on the scores. On the contrary, the SSIM and VIF scores decrease with increasing contrast. Figure 5.2 plots the mean changes in similarity scores over different clip limits used for



(a)

(b)

(c) Ranked processed images

Figure 5.1: An example of quality rankings derived from comparison with a reference image. (a) and (b) show panchromatic images of a sample of the dataset before and after artificial degradation via scraping. The rows of (c) correspond to the different metrics employed; the columns are ordered in ascending quality score. Due to space limitations, only every second column of the ranking is shown.

| | MI | PCC | SSIM | VIF |
|---|---|---|---|---|
| MI | 1.0 | 0.9117 | 0.8189 | 0.7534 |
| PCC | 0.9117 | 1.0 | 0.8004 | 0.7211 |
| SSIM | 0.8189 | 0.8004 | 1.0 | 0.7395 |
| VIF | 0.7534 | 0.7211 | 0.7395 | 1.0 |

Table 5.1: Correlation matrix of different employed similarity metrics, computed over all compared variants.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

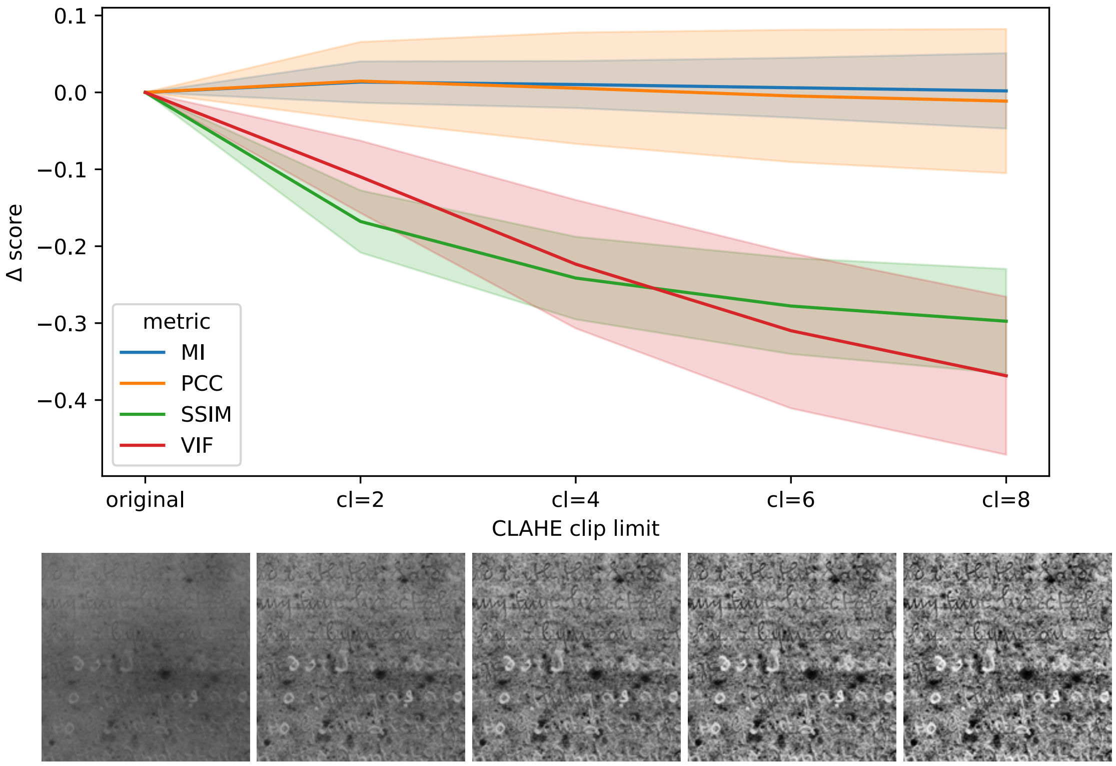


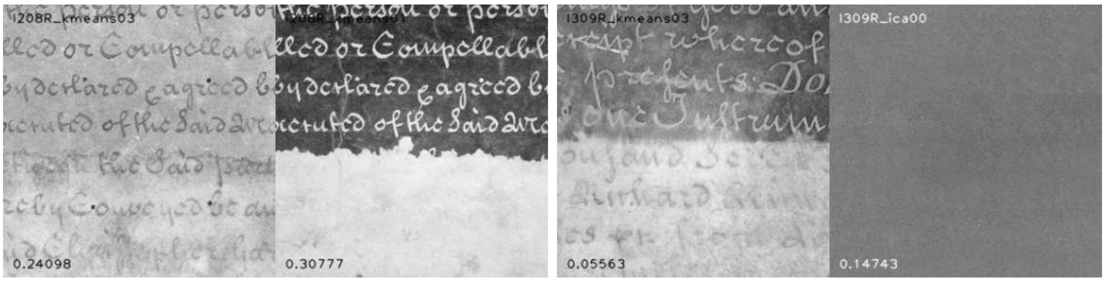
Figure 5.2: The effect of applying CLAHE with increasing clip limits before comparison with the respective metrics. Error bands show the standard deviations. The images below the plot demonstrate the visual effect of contrast enhancements on one example image - note that background structure is enhanced as well as the text.

CLAHE transformations, along with the respective standard deviations. Note that the mean MI and PCC scores remain almost constant, whereby MI exhibits lower standard deviations. MI is thus the most stable of the tested metrics with respect to contrast alterations. The finding that generic contrast enhancements do not improve comparison scores is plausible because the contrast of signal and noise is enhanced likewise. It also suggests that high comparison scores result from contrast that is also present in the original image (especially between text and foreground), which in turn supports the feasibility of image comparison as a quality metric for text restoration.

Although the results are visually convincing in general, individual cases of inexplicable ratings are found frequently; Figure 5.3 shows examples.

### 5.1.5 Discussion

The approach to assess the quality of legibility enhancement methods by comparison with intact reference images is a special case of full-reference IQA. Intuitively the approach is sensible, because the goal of any digital restoration is to produce results as similar to



(a) MI score

(b) PCC score

Figure 5.3: Examples of unexpected ratings. Images on the right were rated higher than images on the left.

the originals as possible. Using four relatively simple image comparison metrics, visually plausible rankings of processed images were produced; however, cases where the method fails were observed as well. In general, the four tested metrics correlate well, with MI and PCC showing the strongest agreement. Also, generic contrast enhancements show no positive effects on the comparison score and MI is identified as the most stable metric in this regard. To definitely validate the feasibility of the approach, a user study is necessary to obtain a strong ground truth dataset containing subjective quality ratings from multiple individuals. For general IQA applications, such datasets are used for the quantitative evaluation of image quality metrics [BCNS18, SWCB05].

The main limitation of this full-reference approach is the dependence on expensively created datasets of artificially degraded manuscripts. In the following sections, no-reference approaches are explored, starting with the description of a new dataset of expert-rated manuscript patches.

## 5.2 The SALAMI Dataset

For the development of general objective IQA methods, the use of public databases containing subjectively rated images is a well-established practice [VNV+15, POMZ+19]. A variety of such datasets have been published; they primarily aim at measuring the perceptual impacts of technical parameters such as image compression artifacts, transmission errors, or sensor quality [SSB06, PLZ+09, PJI+15, VNV+15, LHS19]. However, no such dataset exists for the assessment of text legibility in images.

This section introduces a new dataset of Subjective Assessments of Legibility in Ancient Manuscript Images (SALAMI), designed for the development and validation of objective legibility assessment methods. This dataset consists of 250 images of 50 manuscript regions with corresponding spatial maps of mean legibility and uncertainty, which are based on a study conducted with 20 experts of philology and paleography. As this study is the first of its kind, the validity and reliability of its design and the results obtained are motivated statistically by means of intra- and inter-rater agreement and linear mixed

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

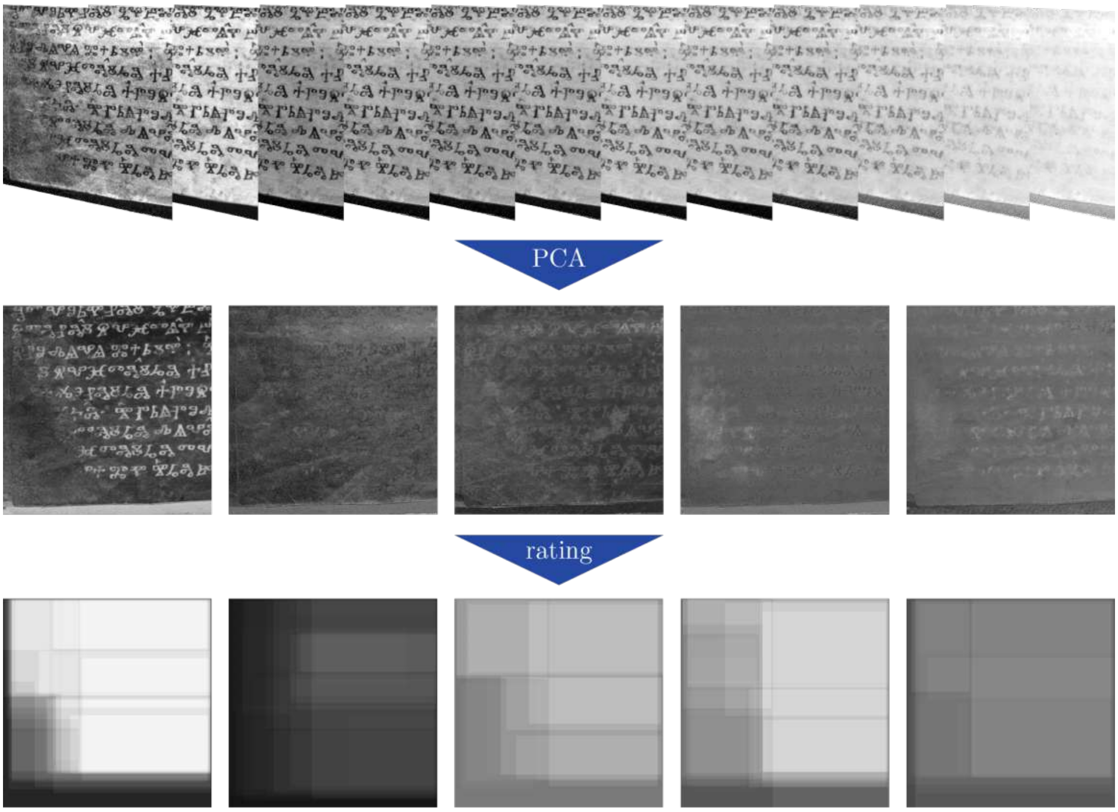


Figure 5.4: Creation of the SALAMI dataset. The original MS image (top) is reduced to 5 principal components, that constitute the test images of the dataset (middle). Mean legibility maps (bottom) are obtained in an expert rating study.

effects models.

### 5.2.1 Study Design

In the following, the design of the subjective IQA study carried out to establish the SALAMI dataset is described and motivated. The documents ITU-T P.910 [Int08] and ITU-R BT.500-13 [Int12] by the International Telecommunication Union provide guidelines for the implementation of subjective IQA studies that are commonly followed in the creation of respective datasets [RFN11, GB16, VNV$^{+}$15, DNT$^{+}$09, LHS19]. In the design of this study, these guidelines are implemented wherever applicable.

#### Test images

The test image set is based on 50 manuscript regions of 60 mm × 60 mm, each of which is represented by 5 images. The regions are sampled from 48 different manuscripts and 8 language families. Slavonic (19), Latin (13) and Greek (12) texts make up the majority of the samples; additionally, two Ottoman texts and one each in Armenian, Georgian,

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

German and Gothic are contained in the dataset. Depending on line height and layout, the regions contain 1-17 lines of text. In the following, the selection of manuscript regions and the creation of the final image set is described in detail, and choices made in the process are motivated.

The manuscript regions represented in the SALAMI dataset are drawn from a set of approximately 4600 pages of 67 historical manuscripts, of which the Computer Vision Lab (TU Wien) has acquired MS images in the course of research projects between 2007 and 2019[1]. The MS images were acquired using different imaging devices and protocols [DS08, HGS12, HBS19], with 6-12 wavebands per image. Rather than presenting whole manuscript pages to the participants, the image contents are reduced to square-shaped regions of 60 mm side-length. This enables the presentation of a fixed region in sufficient magnification, avoiding the need for zooming and panning. The assessment task is thus simplified and accelerated, and the risk of overlooking small pieces of text is minimized. To limit the scope of this study, manuscript regions containing multiple layers of text (such as palimpsests or interlinear glosses) are excluded and must be considered in future work. In order to select 50 suitable regions from the entirety of available pages, a semi-automatic scheme of random selection and human redaction was employed. The scheme ensures that 1) the number of different manuscripts from which the samples are drawn is maximized, 2) the selected regions contain only a single layer of text and 3) the role of a human operator is reduced to deciding if a randomly selected region is suitable for the study (*i.e.*, it contains exactly one layer of text) or not, thus minimizing bias. For the interested reader, further details are provided in the following textbox.

**Selection scheme for manuscript patches.** In a first step, 50 suitable pages are selected: from each of the available manuscripts, a randomly drawn page is presented to an operator, who accepts the image if it contains at least a 60 mm × 60 mm area with exactly one layer of text present (where the text is not required to cover the whole area), and rejects it otherwise. In the next iterations, only manuscripts from which no pages were accepted previously are considered. The process is repeated until (a) a page of each manuscript has been accepted, (b) all pages from the remaining manuscripts have been rejected or (c) the target amount of 50 pages has been reached. In cases (a) and (b) the iteration is restarted, with unvetted pages from all manuscripts available again. In the second step, a similar strategy is used to select the final regions from the drawn images: for each image, randomly selected 60 mm × 60 mm regions are presented to the operator sequentially, until one is accepted.

The source MS images are cropped to the selected square regions and re-sampled to a standard resolution of 12 px/mm (or 304.8 ppi), resulting in an image size of 720 × 720 pixels. This standardization is done to eliminate image size and resolution as a source of variance. For each region, five processed variants are generated to serve as the actual test images. For the sake of simplicity and repeatability, those variants are produced by a principal component analysis on the MS layers[2]. Figure 5.4 shows an example. The

[1] Projects financed by the Austrian Science Fund (FWF) with grant numbers P19608-G12 (2007-2010), P23133 (2011-2014) and P29892 (2017-2019), as well as a project financed by the Austrian Federal Ministry of Science, Research and Economy (2014-2016)

[2] Principal component analysis is frequently used as a standard procedure for dimensionality reduction and source separation in MS manuscript images [ACD18, GCM$^+$17, Min18]

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

inclusion of multiple versions of the same manuscript region enables a versatile use of the SALAMI dataset: additionally to absolute rating applications, it can be used for the development of systems in which relative comparisons between multiple images of the same content are paramount. With 5 variants for each of the 50 manuscript regions, the SALAMI dataset contains 250 test images. According to preliminary tests, this number of images can be assessed in approximately one hour by a single participant.

### Test method

Test methods most commonly used for performing IQA studies are based on degradation ratings, pair comparisons, or absolute ratings [Int08, Int12, POMZ+19, SSB06, MTM12]. Degradation rating approaches assume the existence of an ideal reference image; this is advantageous for experiments in which degradations of an optimal original are evaluated (such as with JPEG compression [SSB06]). In our case, such an optimal reference does not exist, such that this class of methods is not applicable. Pair Comparisons between variants of the same content are shown to provide higher discriminatory power and lower variance than Absolute Ratings, especially when the perceptual differences between those variants are small [PLZ+09, Int08, MTM12]. However, the downsides of this approach are the large number of necessary comparisons ($(n^2 - n)/2$ for $n$ variants) and the lack of a common absolute scale among different contents. The latter problem is solvable when performing cross-content comparisons [POMZ+19]; however, it is not clear if such direct comparisons between different manuscript regions (which vary in preservational condition, size and amount of text and alphabets) are meaningful. For this study, Absolute Ratings are chosen as a base design. Additionally to the above-named reasons, this approach readily allows the following extension for a more detailed specification of legibility: instead of asking the participants to assign a single score to the whole test image, they are required to mark all visible text with rectangular bounding boxes, which are then rated individually (see Figure 5.5). With this approach, a spatial distribution of legibility is obtained instead of a single score per image.

### Rating scale

Following ITU recommendations [Int12], a five-point rating scale is used. As only text legibility and not any other quality of the image should be evaluated, the standard category labels given by ITU ('Excellent', 'Good', 'Fair',...) are refrained from; they could lead to misinterpretations of the task. Instead, the property of legibility is explicitly broken down to the percentage of text within a given area that a participant deems clear enough to read. Dividing the available range into 5 equal intervals leads to the labels and corresponding numerical scores shown in Table 5.2. With the phrasing of legibility in terms of percentages, scores on a true interval scale are obtained, which can not be assumed for the usual qualitative category descriptions [YD13b].

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

| verbal description | numerical value |
|---|---|
| 80-100% readable | 5 |
| 60-80% readable | 4 |
| 40-60% readable | 3 |
| 20-40% readable | 2 |
| 0-20% readable | 1 |
| (non-selected areas) | 0 |

Table 5.2: Available options for legibility ratings.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Bibliothek WIEN Your knowledge hub

### Test environment

Traditionally, subjective IQA experiments are carried out under controlled laboratory conditions [Int08, Int12]. However, Ribeiro *et al.* [RFN11] show that subjective ratings on the LIVE [SSB06] dataset obtained in laboratory conditions can be accurately reproduced in crowd sourcing experiments conducted with Amazon Mechanical Turk. Ghadiyaram and Bovik [GB16] create an IQA dataset of 1162 mobile camera images rated by over 8100 participants online and report excellent internal consistency. Considering these results and the special requirements for the participants (see Section 5.2.1) that lead to a relative shortage of suitable volunteers, the laboratory constraints were loosened, and both participation in controlled conditions and online participation was allowed. A statistical comparison between the results of laboratory and online participation is given in Section 5.2.3. Participants were allowed to ask questions during the instruction and tutorial phases. Online participants could make use of this option via email or phone.

### Order of presentation

The test images are divided into five batches, where each batch contains one variant of each manuscript region. The assignment of the individual variants to the batches is done randomly, but equally for all participants. Within those batches, the images are randomly shuffled for each participant individually. In order to measure intra-rater variability, one randomly chosen image per batch is duplicated.

### Participants

ITU-T P.910 recommends a minimum of 15 participants for any IQA study, while stating that four participants are the absolute minimum that allows a statistical assessment and there is no use in having more than 40 participants [Int08]. For this study, 20 participants were recruited among researchers in the fields of philology and paleography, that have experience in reading original manuscripts. A mixture of university students, pre-doctoral researchers, post-doctoral researchers and professors was aimed at.

### 5.2.2 Experiment Conduction

For an efficient and consistent conduction of the experiment, a web-based user interface is provided for the assessment task, which is equally used by participants in laboratory conditions and online participants. During the primary test, one image at a time is displayed on a medium gray background (50% brightness, following ITU-T P.910 recommendations [Int08]). The participant is required to mark text areas with approximate bounding boxes and individually rate them. For this rating, the estimated amount of legible text within the marked region is chosen from a list, according to Table 5.2. Figure 5.5 shows a screenshot of the primary test interface.

Prior to performing the primary test, the participants must complete three preparatory stages:

1. **Self-assessment.** Participants are required to answer questions about their professional background: academic level, expertise in each of the language families present in the dataset and frequency of exposure to scientific manuscript images (from here on referred to as *SMI exposure*) are queried. The available options for those questions are listed in Table 5.3.

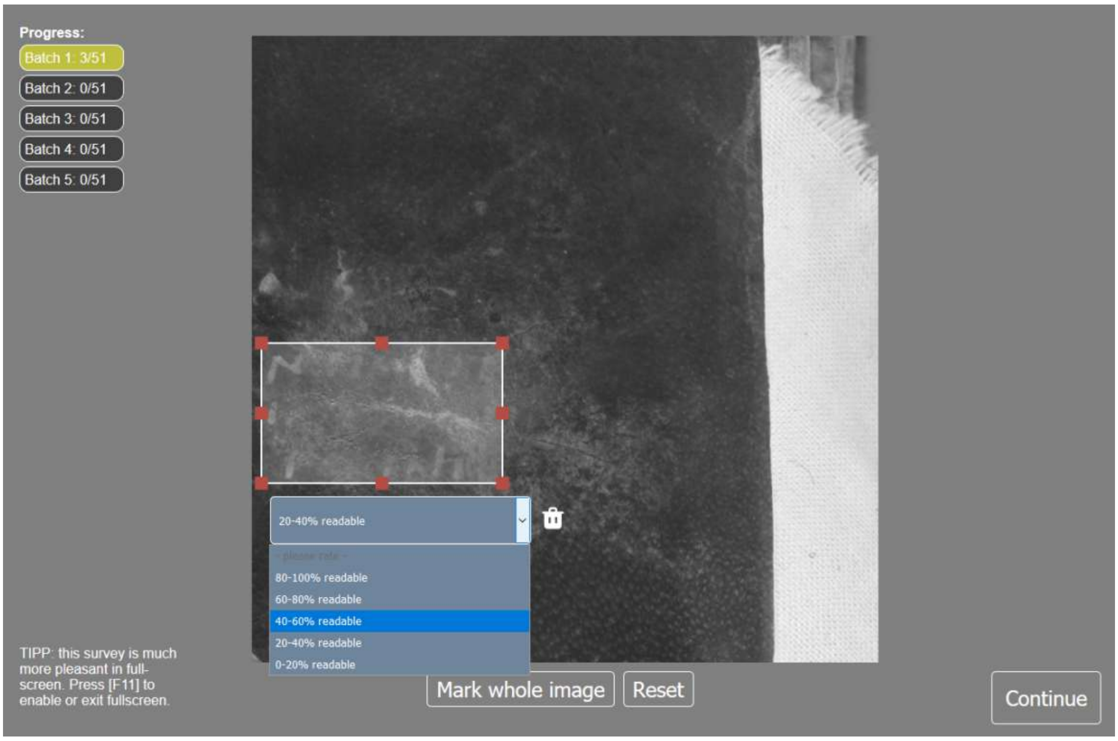
2. **Instructions.** Participants are presented a sequence of pages, in which their task and the functionality of the user interface are explained, each supported by a demonstrative animation.

3. **Tutorial.** 5 images are assessed within the primary test interface without the answers being recorded. These images were manually selected to cover a variety of text coverage and readability levels.

| | **verbal description** | **numerical value** |
|---|---|---|
| language expertise | expert (primary research field) | 3 |
| | advanced (can read and understand) | 2 |
| | basics (knows the script) | 1 |
| | unacquainted | 0 |
| SMI exposure | multiple times a week | 3 |
| | multiple times a month | 2 |
| | occasionally | 1 |
| | never | 0 |
| academic level | professor | 4 |
| | post-doc | 3 |
| | pre-doc | 2 |
| | student | 1 |
| | none | 0 |

Table 5.3: Available options for participant self-assessment.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



TIPP: this survey is much more pleasant in full-screen. Press [F11] to enable or exit fullscreen.

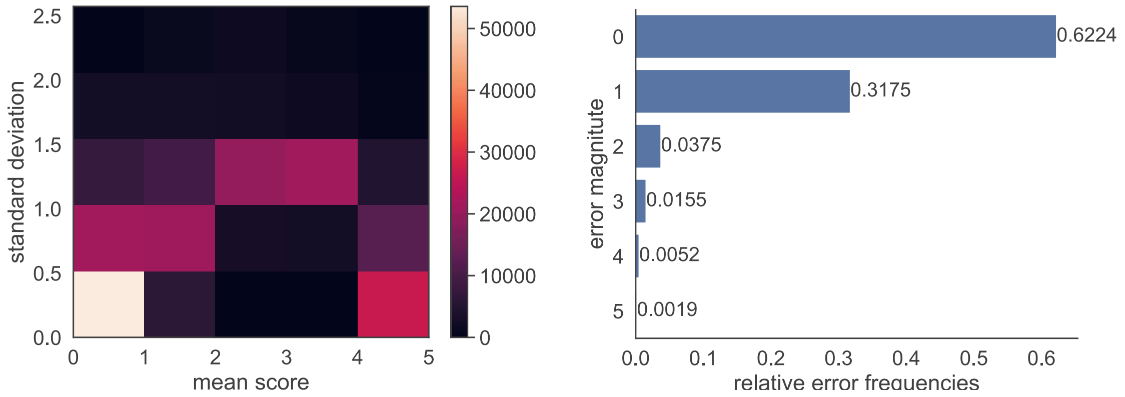
Figure 5.5: A screenshot of the user interface for legibility assessment. The image is displayed on a neutral gray background. The participant is required to mark blocks of visible text with bounding boxes and to estimate how much of this text can be read.

All laboratory test sessions were conducted at the same workplace using a Samsung SyncMaster 2493HM LCD monitor with a screen diagonal of 24 inches at a resolution of 1920x1200 pixels, and a peak luminance of 400cd/m$^2$. Viewing distance was not restricted, as this would not reflect a real situation of manuscript studying. All laboratory participants rated the full set of test images (including duplicates) in a single session. They were allowed to take breaks at any time; however, none of the participants used this option. The online participants used arbitrary monitors and partially took longer breaks (up to days) between their assessments.

### 5.2.3 Evaluation

In total, 4718 assessments were obtained from 20 participants (excluding duplicates for intra-rater variability). Not every participant rated all of the 250 test images, as some of the online participants terminated the test earlier. The median time required to assess a single image was 12.4 seconds (with quartiles $Q_1 = 7.5s$ and $Q_3 = 22.4s$). As the participants were free to select arbitrary image areas for rating, those areas can not be directly related between participants. Instead, each assessment (a given image rated by a given participant) is interpreted as a score map taking on zero in non-selected areas

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



(a) (b)

Figure 5.6: (a) 2D histogram of mean scores and standard deviation of all units. (b) distribution of absolute intra-rater errors of all participants.

and ranging from 1 to 5 in selected areas, according to their rating. Intersection areas of overlapping bounding boxes receive their maximum score. The choice of the maximum (over a median or rounded mean) was motivated by the observation that laboratory participants deliberately placing one bounding box on top of another always intended to label small areas with higher legibility than their surroundings, and never the other way around. For the following analysis, the images are partitioned into observational *units* of 2 mm × 2 mm (corresponding to 24 × 24 pixels). An elementary *observation* is defined as the rounded mean of scores assigned to the pixels of a given unit by a single participant; accordingly, an image assessed by one participant results in 900 observations. A first look at the relative frequencies of scores over all observations leads to the insight that the bulk of observations report background, *i.e.*, units where no text is visible. Furthermore, the 2D histogram of unit-wise mean scores and standard deviations (Figure 5.6a) shows a concentration of units with low scores and low standard deviations. This suggests that participants largely agree on the distinction between foreground and background. In order to prevent this mass of trivial background observations from biasing the analysis of intra- and inter-participant agreement, units that are labeled as background by more than half of the participants are excluded from all statistical considerations of this section.

### Participant characteristics and agreement

Participant screening was performed following the algorithm described in ITU-R BT.500 [Int12]. None of the participants were rejected. Table 5.4 shows the self-assessed properties of all 20 participants, where the numbers refer to the verbal descriptions given in Table 5.3.

**Intra- and inter-rater variability** To assess intra-rater variability and thus the repeatability of the experiment, the absolute errors between units of duplicate presentations (see Section 5.2.1) are recorded. In accordance with the evaluation strategy given at the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

| *participant ID* | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *academic level* | 2 | 2 | 2 | 2 | 4 | 2 | 4 | 4 | 2 | 1 | 1 | 2 | 4 | 4 | 3 | 4 | 3 | 3 | 4 | 3 |
| *SMI exposure* | 3 | 3 | 3 | 1 | 3 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 1 | 3 | 3 | 2 | 2 | 3 |
| *Armenian* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Georgian* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *German* | 3 | 3 | 2 | 2 | 2 | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 0 | 1 | 3 | 2 | 3 |
| *Gothic* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Greek* | 1 | 0 | 2 | 0 | 1 | 3 | 1 | 2 | 0 | 0 | 1 | 0 | 2 | 2 | 1 | 3 | 3 | 1 | 3 | 2 |
| *Latin* | 2 | 1 | 2 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 3 | 2 | 3 |
| *Ottoman* | 0 | 0 | 3 | 3 | 3 | 0 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Slavonic* | 0 | 0 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 3 | 1 | 0 | 0 | 3 | 0 |

Table 5.4: Participants overview. Verbal descriptions of the numerical values are given in Table 5.3.

| | ICC(2,1) | ICC(3,1) | ICC(2,k) | ICC(3,k) |
|---|---|---|---|---|
| *all participants* | 0.668 [0.64, 0.69] | 0.711 [0.71, 0.71] | 0.976 [0.97, 0.98] | 0.976 [0.97, 0.98] |
| *lab participants* | 0.702 [0.65, 0.74] | | | |
| *online participants* | 0.664 [0.63, 0.69] | | | |
| *SMI exposure $>$ 1x/week* | 0.690 [0.66, 0.72] | | | |
| *SMI exposure $\leq$ 1x/week* | 0.649 [0.57, 0.71] | | | |

Table 5.5: ICCs with 95% confidence intervals, for different participant groups.

beginning of this section, only units with a score greater than zero in at least one of the duplicate presentations are considered. The distributions of absolute errors is shown in Figure 5.6b. The mean absolute error across all duplicate observations is **0.469**.

The agreement of different raters (participants) on legibility scores was measured using Intraclass Correlation Coefficients (ICC). Following the definitions of Shrout and Fleiss [SF79], the ICC variants ICC(2,1), ICC(3,1), ICC(2,k) and ICC(3,k) are reported. While the variants ICC(2,_) view the set of participants as a random sample from a larger population and thus express the reliability of the proposed experimental design, variants ICC(3,_) express the reliability of the specific dataset that is published, rated by the specific participants of this study. On the other hand, ICC(_,1) estimate the reliability of a single participant, while ICC(_,k) estimate the reliability of the average of k (in our case k=20) participants [SF79]. The results are shown in Table 5.5, along with their 95% confidence intervals.

**Impact of the test environment and expertise** As the study was conducted as a mixture of 7 laboratory sessions and 13 online sessions, it is worth assessing the influence of the test environment on rater agreement; uncontrolled factors in the online environment (such as monitor characteristics or insufficient understanding of the task) could lead to greater divergence between participants. Thus, ICC(2,1) is reported for each of the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

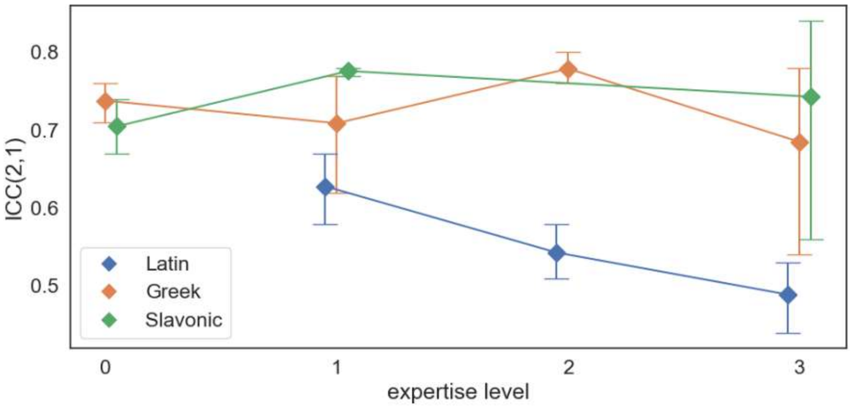


Figure 5.7: Inter-rater agreement for different levels of expertise in Latin, Greek and Slavonic. Vertical bars show 95% confidence intervals. Missing values are due to an insufficient number of participants in the respective expertise category.

groups separately; other ICC variants are omitted, as this question shall be addressed independently of a specific set/number of participants. As shown in Table 5.5, the ICC estimate is indeed higher for the lab participants; however, note the large confidence intervals, where the ICC estimate for online participants is within the 95% confidence interval of the lab participants.

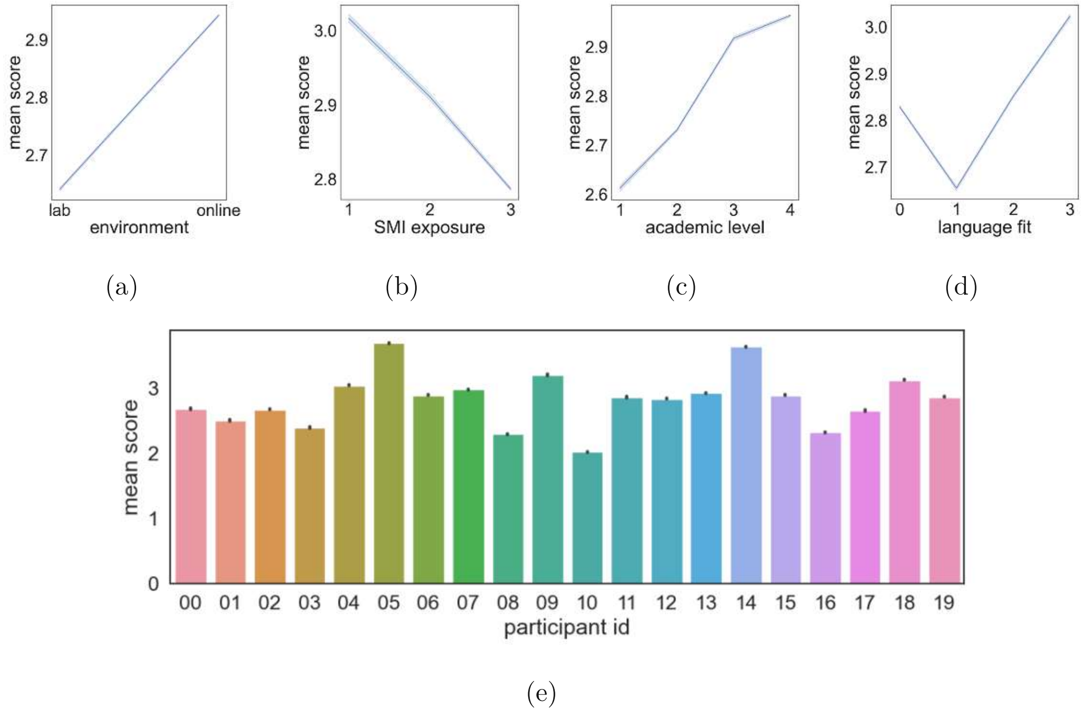
One hypothesis related to participant expertise is that increased professional exposure to scientific manuscript imagery improves inter-rater agreement. Again, ICC(2,1) is computed separately for participants who work with such imagery multiple times a week and all participants with less experience (a finer distinction is refrained from, due to a small number of participants in the lower exposure categories). The results are shown at the bottom of Table 5.5. As with the test environment, an effect on the ICC can be observed, but with largely overlapping confidence intervals.

Further, the influence of language expertise on agreement is investigated. For this, the observations on images containing Latin, Greek and Slavonic (which constitute the bulk of the test images) are treated separately. For each of those image sets, observations are partitioned according to the expertise level of their raters in the respective language and compute ICC(2,1) scores. The resulting estimates, along with their 95% confidence intervals, are visualized in Figure 5.7, where no general trend is observable.

### Systematic effects and sources of variation

After addressing the effects of participant and experimental parameters on inter-rater agreement, we now turn to the analysis of their systematic effects, *i.e.*, their influences on mean scores. Furthermore, these effects are compared to uncontrolled variations between participants and, most importantly, to the variation introduced by the properties of the observed units. The following controlled parameters which gave reason to suspect linear

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.



(a) (b) (c) (d)

(e)

Figure 5.8: Effects of participant parameters on mean scores: (a) test environment, (b) exposure to scientific manuscript images, (c) academic level, (d) language fit, *i.e.*, for a given observation, the expertise of the participant in the language of the observed manuscript patch. In (e) the mean score for each participant is shown.

relations (see Figures 5.8a-5.8d) are considered: test environment, SMI exposure, academic level and *language fit*. The language fit of an observation is defined as the participant's skill level in the language associated with the observed unit. For the possible values of the above-mentioned parameters refer to Table 5.3. The sum of those influences plus an uncontrolled (random) source of variation constitutes the participant-wise variation of means shown in Figure 5.8e.

In order to jointly model and investigate those participant-specific effects and the effects of observed units, linear mixed effects models are employed. In the full model, legibility score is the dependent variable. Fixed effects are test environment, SMI experience, academic level and language fit. As further uncontrolled variations between participants are to be expected, the participant ID is included as a random intercept. A second random intercept is defined as the ID of the observed unit to model the dependence of an observation's score on the observed unit. The model was fitted in *R* [R D08] using the *lme4* package [BMBW15]. For each of the fixed effects, a likelihood ratio test of the full model against a model without the respective effect was performed. Language fit has been found to affect legibility scores by an increase of **0.05** per skill level, at a $p < 0.001$ confidence level. For the other fixed effects, p values are above 0.05. The random effects

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

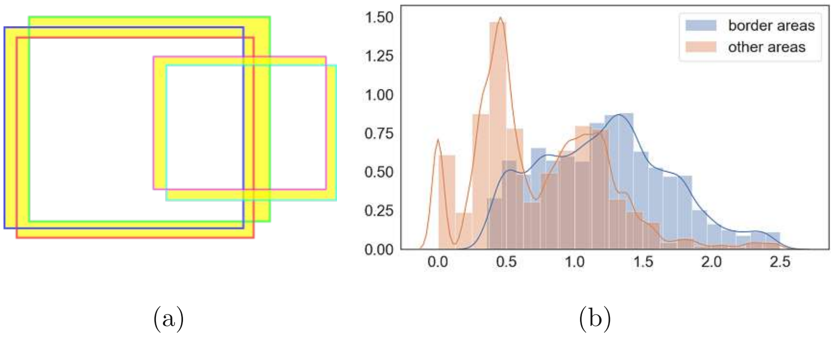


Figure 5.9: (a) Definition of critical border areas. The differently colored rectangles are bounding boxes defined by different participants. The union of pairwise symmetric differences of similar bounding boxes (in terms of intersection over union) is shown in yellow. (b) Distribution of standard deviations for border areas versus other areas.

'unit ID' and 'participant ID' contribute a variance of **2.133** and **0.136**, respectively; the residual variance is **0.803**.

### Spatial distribution of variability

Finally, patterns in the spatial distribution of units with a high variability in legibility scores are explored. Qualitative inspection of standard deviation maps that are part of the published dataset (see Figure 5.10c for an example) suggests that the highest variability is found near the boundaries of text areas and is thus caused by the variations in bounding box placement. To test this hypothesis, such critical border areas are defined as the union of symmetric differences between pairs of similar bounding boxes from different participants; in this context, bounding boxes are considered similar if their intersection-over-union ratio is greater than 0.9. The idea is illustrated in Figure 5.9a. It was found that 7.56% of units fall under the definition of border areas given above. The distributions of standard deviations of border areas and other areas are shown in Figure 5.9b. The mean standard deviation of border areas (**1.218**) is significantly higher than the mean standard deviation of other areas (**0.750**) at a $p < 0.001$ confidence level (according to a Mann-Whitney U test as the respective distributions are non-normal).

## 5.2.4 Dataset description and validity

As described in Section 5.2.1, the SALAMI dataset is based on 250 test images: 50 manuscript regions are represented by 5 processed variants each. Along with every test image, a legibility map averaging the scores of all participants is published, as well as a standard deviation map showing the spatial distribution of uncertainty. See Figure 5.10 for an example of such an image triplet. The test images and their legibility maps are ready to be used as a ground truth for developing computer vision methods for the localized estimation of legibility in manuscript images. Additionally to the estimation of absolute

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

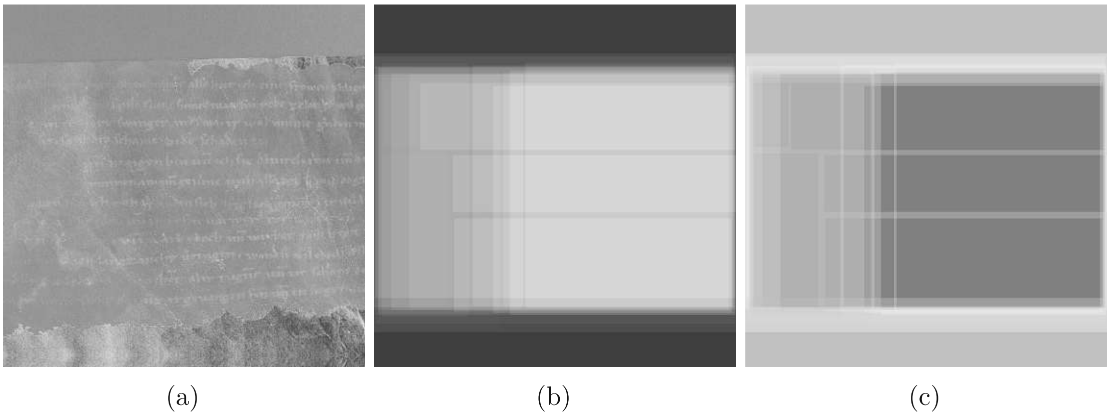


Figure 5.10: An instance of the dataset, consisting of the test image (a) and corresponding mean score map (b) and standard deviation map (c).

legibility from a single image, the dataset also supports pairwise comparison or ranking applications due to the five variants per manuscript region contained in the dataset. The standard deviation maps can be used to exclude regions with a high uncertainty from training/evaluation, or for implementing a weighted loss function depending on local uncertainty. Additionally to the pixel maps, the originally recorded data (.json encoded) are provided, along with documented Python scripts that can be used to reproduce the legibility and score maps as well as the results described in Section 5.2.3. The dataset, code and documentation are available on **zenodo** [Bre20]. As the SALAMI dataset is the first of its kind, its validity, as well as the appropriateness of the novel study design used to generate it, must be justified. For this purpose, the results of the statistical analyses obtained in Section 5.2.3 are summarized (for interpretation of numerical values provided, remember that they are related to rating scores in the interval $[0, 5]$).

**Repeatability:** Even with background areas excluded, 62.2% of absolute intra-rater errors are zero and the mean absolute error is 0.469.

**Reliability:** Table 5.5 summarizes the inter-rater agreement and thus reliability of the test method (ICC(2,_)) as well as the specific test results published in the form of the dataset (ICC(3,_)). According to the highly cited guidelines by Koo *et al.* [KL16], in either case moderate reliability can be expected from a single rater (ICC(_,1)) but excellent reliability from the average of 20 independent raters (ICC(_,k)). No conclusive evidence for a negative impact of online participation or lack of participant skills/experience on reliability has been found.

**Systematic effects:** The linear mixed effects model analysis does not show a statistically significant impact of SMI experience, test environment or academic level on the scores obtained. Only the property of language fit has a significant effect; however, the practical relevance of an increase in mean scores by 0.05 for each language skill level is negligible. Furthermore, it should be noted that the language skills of participants were acquired through self-assessment and not validated in any way.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

**Sources of variation:** The second and most important conclusion that can be drawn from the linear mixed effects model is that the identity of the observed unit contributes the majority of variance in the scores (2.133), while the identity of the participant only contributes a variance of 0.136. This is an indicator that the measured legibility is an objective property of a given image area and not dominated by subjective preferences of the participants.

**Spatial distribution of uncertainty:** Uncertainty is varying spatially within test images and especially border regions of text areas exhibit an increased variability in scores. This motivates the publishing of standard deviation maps along with mean score maps as an essential part of the dataset.

### 5.2.5 Discussion

A study with 20 experts of philology and paleography was conducted to create the first dataset of Subjective Assessments of Legibility in Ancient Manuscript Images, intended to serve as a ground truth for developing and validating computer vision-based methods for quantitative legibility assessment. Such methods, in turn, could elevate a whole research field centered around the digital restoration of written heritage. Additionally to creating the dataset itself, a novel methodology to conduct similar studies in the future is described, demonstrating the validity of the results and the robustness against variations in test environment and participant properties. The dataset can be directly used as a ground truth for legibility assessment methods that aim at estimating a spatial distribution of absolute legibility over an image, for example, using the correlation between estimated scores and mean scores from the dataset as a performance metric. Furthermore, it is applicable for the evaluation of ranking methods: with five variants per manuscript region, direct comparisons between the legibility maps of those variants are possible (both locally and globally averaged across the whole image).

Qualitative comments of the participating experts regarding their perspective on the study design were collected, which revealed potential for improvement in similar studies carried out in the future. Specifically, the following issues were pointed out:

- Line height has an impact on legibility. This property of the assessed images is not considered in the analysis.

- The study only assesses the percentage of text that is readable; however, another relevant dimension would be the time/effort required to read a text.

- Manuscript experts tend to dynamically 'play' with the images (*i.e.*, vary contrast, brightness, scale, *etc.*) in order to decipher texts. This was not permitted in the test design, potentially biasing the results.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

## 5.3 A baseline for estimating human legibility in images of degraded text

The lack of established methods to quantitatively assess document image quality or the success of text restoration methods can be attributed partly to the previous absence of a suitable ground truth dataset of subjectively rated manuscript images, the like of which is usually employed for the development and validation of general IQA methods [VNV$^+$15, SSB06, PJI$^+$15]. After having introduced such a dataset in the previous section, its use for the evaluation of methods for legibility estimation is now demonstrated.

Table 5.6 summarizes the relevant approaches reviewed in Section 2.3. All of the approaches in the full/reduced reference category were previously used to estimate handwritten document images or evaluate restoration approaches; however, none of them was shown to correlate to human legibility. Furthermore, the dependence on reference information limits the applicability and thus contradicts Requirement **R3** formulated at the beginning of this chapter. In the no-reference category, only the approaches that operate on binary images are disqualified (**R4**). The other approaches are in principle valid candidates for a legibility estimator; yet their correspondence to human legibility and robustness to script and language remains to be shown.

In the following, a baseline for the problem of estimating human legibility in images is established by defining an evaluation framework and testing candidate legibility estimators for their correlation with human perception.

### 5.3.1 Experiments

This section describes the experiments conducted to test the correlation of candidate legibility estimators with human legibility, using the SALAMI dataset (Section 5.2) as a ground truth. After defining the test framework and tested candidate methods, experimental results are described.

#### Test framework

The experiments aim at testing the outputs of candidate legibility estimators for rank correlation with the subjective human legibility scores of the SALAMI dataset, which are reported in the form of Spearman Rank Correlation coefficients (SRC) [PNR15]. Evaluation takes place with respect to the following dimensions:

**Spatial accuracy.** The correlations are evaluated on several spatial levels: The rating of the whole image (720 × 720 pixels or 60 mm × 60 mm), as well as the rating of sub-squares of 360, 180, 90 and 45 pixels side-length are evaluated. The ground truth legibility scores are obtained by averaging the legibility maps of the SALAMI dataset across the corresponding regions.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

| *description* | *publication* | *reference* | *limited applicability* |
|---|---|---|---|
| **full/reduced reference approaches** | | | |
| comparison to non-degraded reference | Giacometti *et al.* 2017 [GCM+17] | image of the object before artificial degradation | dedicated datasets only |
| cluster separability | Arsene *et al.* 2018 [ACD18] | pixel labels | - |
| potential contrast | Shaus *et al.* 2017 [SFGST17] | pixel labels | - |
| comparison to ground truth binarization | multiple (see *e.g.* [SON19, PZK+19]) | ground truth binarization | binarized documents |
| OCR performance | multiple (*e.g.* [LSDS11, HDS14]) | ground truth transcription | known alphabets / languages |
| **no-reference/blind approaches** | | | |
| General blind IQA | multiple (*e.g.* [MMB12, YKKD13, BMM+18] | human perceived quality | - |
| Blind document IQA | multiple (*e.g.* [YD13a, LZQ18, LZQ19, LD19]) | OCR performance on machine written documents | - |
| handcrafted features + SVM | Stommel & Frieder 2011 [SF11] | human legibility rating (good, medium, bad) | binarized documents |
| handcrafted features + NN | Obafemi & Agam 2012 [OAA12] | human quality rating | binarized and segmented characters |
| gabor filter responses + support vector regression | Shahkolaei *et al.* 2018 [SNAMC18] | human quality rating | - |
| Text detection | multiple (*e.g.* [NM12, LSB+17, ZYW+17]) | labeled text areas | - |

Table 5.6: Overview of related work and approaches to legibility estimation.

**Absolute legibility vs. single content ranking.** The ability to estimate absolute legibility is measured by correlating the estimates with the ground truth across all manuscript regions. As the SALAMI dataset contains five image versions for each manuscript region, it can also be used to test the ability of an estimator to correctly rank different versions of the same content. This is done by computing the SRC separately for each manuscript region. This per-content correlation is a weaker property than the overall correlation (as the former is implied by the latter); however, it is sufficient for applications like the comparison of different enhancement algorithms and automated parameter optimization.

**All areas vs. text areas only.** Lastly, the performance on the whole image is compared with the performance on text areas only. These test cases correspond to different application scenarios and pose different challenges: if no additional information about the content is available, the estimators must be able to handle background areas and assign minimal legibility to them. On the other hand, if the location of text is known *a priori*, the estimator must not handle background areas and only the discriminatory power between text legibility levels is relevant. Within the SALAMI dataset, text areas are defined as areas with a mean legibility score of at least 1 in at least one of the 5 image variants; following this definition, text areas make up ca. 80% of the dataset.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Some of the evaluated methods are trained on the SALAMI dataset. In these cases, a leave-one-out cross-validation approach is applied: For each of the 50 manuscript regions represented in the dataset, an own model is trained, with the 5 corresponding images excluded from training and used for testing; the reported scores are the means of all trained models. This approach is chosen in order to cope with a relatively small dataset.

### Tested candidate estimators

The SALAMI dataset consists of input images paired with spatial maps of subjective legibility, without additional information. Therefore, only blind/no-reference candidate estimators (see Section 2.3.2) can be tested within this framework. In the following, the specific metrics and methods that are tested for correlation with subjective legibility scores are described.

**Simple image statistics.** For a basic reference, the correlation of legibility scores to image statistics with plausible impact on legibility is tested:

- *RMS contrast:* Equivalent to the standard deviation of intensities: $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$, where $n$ is the number of pixels in the area of interest, $x_i$ is the intensity of pixel $i$ and $\mu$ is the mean intensity.
- *entropy:* Shannon entropy of intensities: $-\sum_{i=1}^{n}\mathrm{P}(x_i)\log_2\mathrm{P}(x_i)$, where $n$ is the number of pixels in the area of interest and $\mathrm{P}(x_i)$ is the probability of pixel $i$ having the intensity $x_i$, based on a histogram of intensities.
- *edge intensity:* Edge intensity images are computed as magnitudes of gradients resulting from convolution with horizontal and vertical Sobel operators. The edge intensity images are then averaged over the area of interest.

**Text detection and layout analysis.** This class of methods detects text in natural scenes or documents. The amount of detected text lines in a given image area (possibly weighted by corresponding detection confidences) could be an implicit indicator for human legibility. To obtain legibility scores from detected text regions, first a *detection map* is created: depending on the specific method, pixels inside of detected regions are set to the associated detection confidences or to one, if no detection confidence is available; background pixels are set to zero. The legibility score of a given area of interest is then defined as the mean of the corresponding detection map. None of the methods were trained on the SALAMI dataset. The following approaches are tested:

- **Neumann & Matas 2012** [NM12]: A scene text detector based on cascade classification of extremal regions, trained on segmented characters and non-character images. Binary detection maps of multiple input layers [NM12] are summed up.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

- **Zhou *et al.* 2017 (EAST)** [ZYW+17]: A scene text detector based on a convolutional neural network (CNN), trained on the ICDAR 2015 dataset [KGBN+15]. In the resulting detection map, regions are weighted by their confidences.
- **Liao *et al.* 2017 (TextBoxes)** [LSB+17]: A CNN-based scene text detector trained on the ICDAR 2013 dataset [KSU+13]. In the resulting detection map, regions are weighted by their confidences.
- **Grüning *et al.* 2019** [GLS+19]: A CNN-based text line detector for layout analysis, as implemented in the Transkribus plattform [KCHM17]. Evaluation is done both on text line level and text region level. Detection maps are binary.

**(Document) Image Quality Assessment.** Blind IQA methods estimate quality scores for a given input image, which can be directly used as legibility estimates. Generic "image quality", especially with respect to natural images, might prioritize image properties irrelevant for human legibility. The selected methods are thus evaluated once trained on the datasets used in the original publications, and once trained on the SALAMI dataset with its legibility scores. The following approaches are tested:

- *Mittal et al. 2012 (BRISQUE)* [MMB12]: A popular general IQA metric appearing as a reference implementation in document IQA approaches [SNAMC18, GC16]. The original version is trained on the LIVE dataset [SSB06].
- *Shahkolaei et al. 2018* [SNAMC18]: A dedicated document IQA method for historic manuscript images. The original version is trained on the authors' own dataset [SNAMC18].

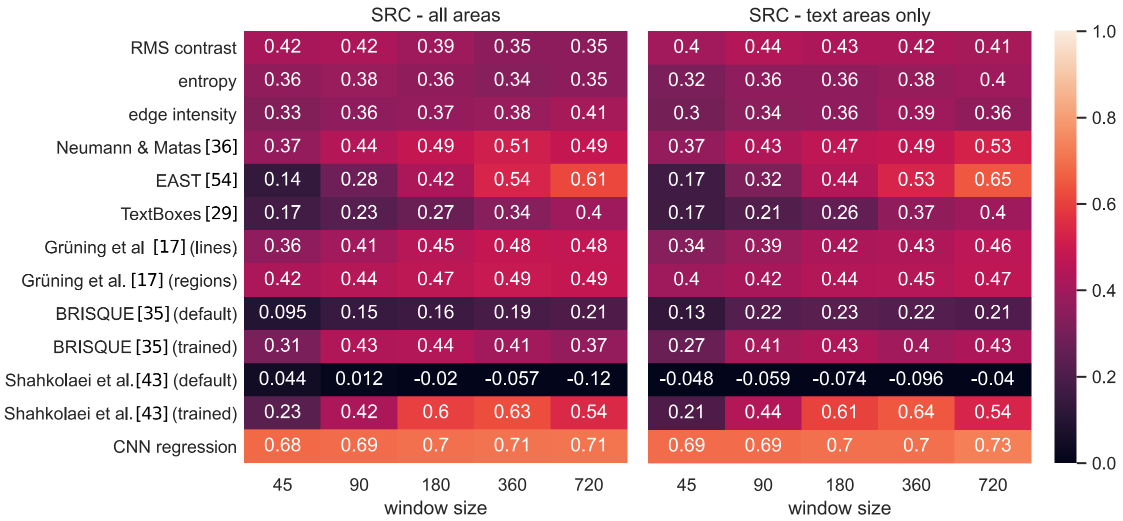
**A dedicated CNN.** Finally, a CNN is trained for legibility estimation from scratch, in order to obtain a baseline for dedicated deep-learning based legibility estimation:

- *CNN regression:* The network architecture is based on an 18-layer ResNet [HZRS16], equipped with an extra linear layer for regression. Considering the leave-one-out cross-validation approach mentioned at the beginning of the section, 50 models are trained for 20 epochs each on 240 × 240 pixel patches of images of the SALAMI dataset. For both training and inference, the patches are extracted from the original images in a sliding window manner with a stride of 10 pixels. Inference results are thus directly interpretable as legibility maps compatible with the evaluation framework.

### 5.3.2 Results

Figure 5.11 shows the SRCs of tested candidate estimators with the legibility scores of the SALAMI dataset for different spatial resolution levels (window sizes). The performance of most of the tested candidates is comparable to the performance of simple image statistics

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub



| | SRC - all areas | | | | | SRC - text areas only | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| window size | 45 | 90 | 180 | 360 | 720 | 45 | 90 | 180 | 360 | 720 |
| RMS contrast | 0.42 | 0.42 | 0.39 | 0.35 | 0.35 | 0.4 | 0.44 | 0.43 | 0.42 | 0.41 |
| entropy | 0.36 | 0.38 | 0.36 | 0.34 | 0.35 | 0.32 | 0.36 | 0.36 | 0.38 | 0.4 |
| edge intensity | 0.33 | 0.36 | 0.37 | 0.38 | 0.41 | 0.3 | 0.34 | 0.36 | 0.39 | 0.36 |
| Neumann & Matas [36] | 0.37 | 0.44 | 0.49 | 0.51 | 0.49 | 0.37 | 0.43 | 0.47 | 0.49 | 0.53 |
| EAST [54] | 0.14 | 0.28 | 0.42 | 0.54 | 0.61 | 0.17 | 0.32 | 0.44 | 0.53 | 0.65 |
| TextBoxes [29] | 0.17 | 0.23 | 0.27 | 0.34 | 0.4 | 0.17 | 0.21 | 0.26 | 0.37 | 0.4 |
| Grüning et al [17] (lines) | 0.36 | 0.41 | 0.45 | 0.48 | 0.48 | 0.34 | 0.39 | 0.42 | 0.43 | 0.46 |
| Grüning et al. [17] (regions) | 0.42 | 0.44 | 0.47 | 0.49 | 0.49 | 0.4 | 0.42 | 0.44 | 0.45 | 0.47 |
| BRISQUE [35] (default) | 0.095 | 0.15 | 0.16 | 0.19 | 0.21 | 0.13 | 0.22 | 0.23 | 0.22 | 0.21 |
| BRISQUE [35] (trained) | 0.31 | 0.43 | 0.44 | 0.41 | 0.37 | 0.27 | 0.41 | 0.43 | 0.4 | 0.43 |
| Shahkolaei et al. [43] (default) | 0.044 | 0.012 | -0.02 | -0.057 | -0.12 | -0.048 | -0.059 | -0.074 | -0.096 | -0.04 |
| Shahkolaei et al. [43] (trained) | 0.23 | 0.42 | 0.6 | 0.63 | 0.54 | 0.21 | 0.44 | 0.61 | 0.64 | 0.54 |
| CNN regression | 0.68 | 0.69 | 0.7 | 0.71 | 0.71 | 0.69 | 0.69 | 0.7 | 0.7 | 0.73 |

Figure 5.11: SRCs between tested methods and ground truth (SALAMI dataset [BS21]), given separately for different window sizes. Results for the entire dataset are shown on the left, results for only text areas on the right.

such as the *RMS contrast*. Exceptions are the dedicated *CNN regression*, as well as the method by Shahkolaei *et al.* [SNAMC18] when trained on the SALAMI dataset. The EAST scene text detector [ZYW$^+$17] out of the box performs reasonably well on large images, without any training on the SALAMI dataset. No statistically significant differences were observable when comparing the results for the entire dataset with the results for text areas only.

Results of the single content ranking tests (*i.e.*, the SRCs are computed for each manuscript region separately, as described in Section 5.3.1) are shown as box plots in Figure 5.12. From this experiment, additional insights are gained:

1. The interquartile ranges indicate high variability in performance between different manuscript regions. The *CNN regression*, which shows the best overall correlation, also appears to be most invariant to the input region.

2. Comparing the medians to the overall SRCs (Figure 5.11), no general trend is observable: some of the tested methods show higher correlations in the single content situation, while others show lower correlations. Notable in this context are the results of *RMS contrast* and *entropy* on the largest window size, where they are only surpassed by *Neumann & Matas* (which is also performing unusually well here) and the *CNN regression*.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

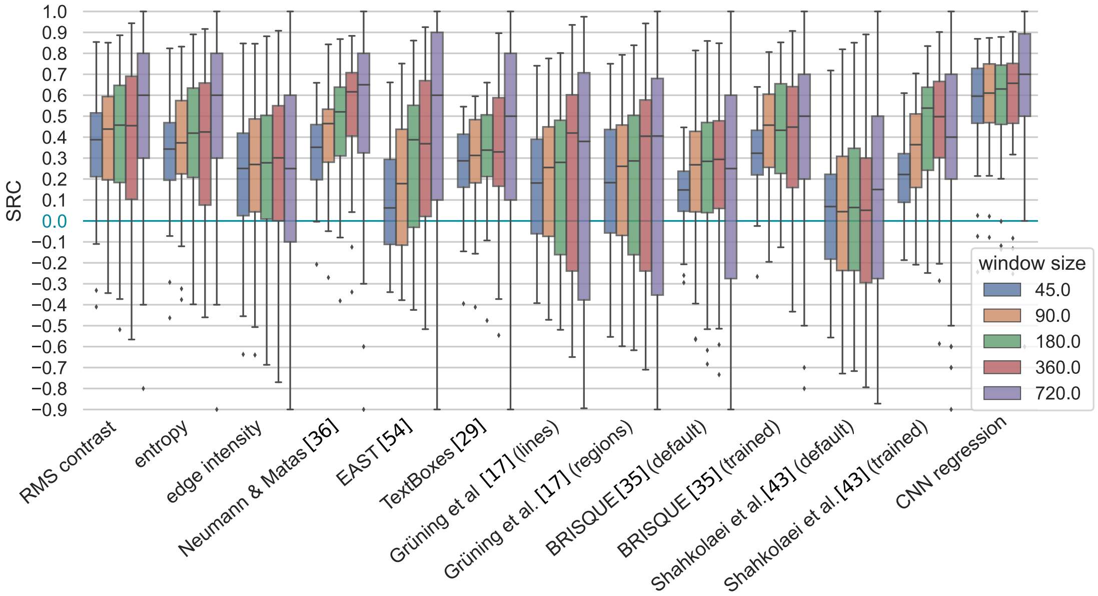


Figure 5.12: Boxplots of SRCs computed for each manuscript region separately, grouped by different window sizes. The boxplots show the quartiles of SRCs observed in the different regions, outliers are shown as dots.

### 5.3.3 Discussion

Reviewing previous approaches to evaluate the quality of manuscript images or the success of text restoration methods, no metric suitable for this purpose is found: either the approaches have a limited applicability (specialized dataset, binary images, machine-written documents) or they are insufficiently validated with respect to domain expert ratings. In order to establish a baseline for further research in this area, an evaluation framework for legibility estimators that is based on correlation with the subjective legibility scores of the novel SALAMI dataset [BS21] is proposed. Several possible candidates for legibility estimators are then evaluated based on this framework.

In the experiments conducted, few of the tested methods surpass basic image statistics such as *RMS contrast*. The work of Shahkolaei *et al.* [SNAMC18] is conceptually most compatible with the definition of a valid legibility metric given at the beginning of this chapter. Interestingly, the results are close to random when applied to the SALAMI dataset in its originally trained form; when trained on the SALAMI dataset, however, the performance increases drastically. This behavior might result from the fundamental differences in the study designs used to create their dataset and the SALAMI dataset (pair comparisons of whole pages by technical students vs. spatially resolved absolute ratings by domain experts). Also the *BRISQUE* [MMB12] IQA metric shows a significant gain in correlation after training on the SALAMI dataset, suggesting that generic image quality does not directly translate to human legibility. However, even after training, the

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

performance stays in the range of simple image statistics.

The legibility estimates generated via text detection and layout analysis show mediocre performance. However, the translation from text detections to legibility scores adopted for this study is rather crude and is worth further elaboration in future work.

The best results are obtained with a CNN regression trained from scratch on the SALAMI dataset. The problem with using CNN predictions as a quality metric, however, lies in the lack of decision transparency [CSHB18]; when pursuing a CNN path in future work, this aspect must be given the same attention as the improvement of correlation to human perception.

## 5.4 Summary

When developing estimators for the quality of graphical heritage imagery, full-reference IQA is a stable option, but only applicable with expensively created datasets of artificially degraded objects. To enable the development of more flexible no-reference IQA approaches, a novel dataset of expert-rated manuscript images, created with a novel study design, is introduced. The validity of the ratings obtained is confirmed by statistical evaluations. Using the dataset as a ground truth, several candidate quality estimators are evaluated, with a CNN-based regression achieving the best correlation to human ratings.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

# CHAPTER 6

# Conclusion

We are twelve billion light years
from the edge. That's a guess.

*Katie Melua*

This work is dedicated to recording and visualizing graphical heritage that manifests in variations of reflectance or geometry, in order to improve its accessibility in humanist scholarship. The specific research aims formulated in favor of this high-level aspiration are concerned with optimizing methods of multi-light imaging for the purpose and finding approaches for objectively evaluating the resulting visualizations. In the following, the contributions described in Chapters 3-5 are assessed with respect to these aims, thereby addressing strengths, weaknesses and open questions.

### Optimizing multi-light imaging for the acquisition and visualization of graphical heritage

The assembly of a mobile acquisition system for Multispectral (MS) imaging and Photometric Stereo (PS) described in Section 3.1 does not advance multi-light imaging methodically, but in terms of availability and practicality. It provided a solid basis for subsequent investigations and for the implementation of research projects involving numerous on-site imaging campaigns. A feature not seen in any commercial or academic solution is the method for calibration and automated correction of longitudinal chromatic aberrations ('focus shifts') described in Section 3.2. While the approach is capable to compensate for focus shifts occurring in any existing imaging system, it also introduces additional complexity. Thus, its usefulness for a practical application depends on the color correction capabilities already provided by the lens used and the application-specific requirements on image sharpness. The practice-proven heuristics for PS light source calibration are, on the one hand, part of the documentation of the datasets and experiments described in Chapter 4; on the other hand, they are provided with the hope of being useful for practitioners, as parts the implementations described are not found in literature.

Chapter 4 investigates general questions in PS, that are especially relevant for graphical heritage applications. The contributions on optimal light source configurations for quasi-flat surfaces, while certainly applicable to other application domains, are conducted with specific requirements in mind: first, carrier surfaces of graphical heritage tend to be quasi-flat; second, equipment size and acquisition time are non-negligible parameters for the feasibility of on-site imaging campaigns with spatial and temporal constraints. Thus, this work has contributed to an optimization of acquisition processes by showing that the reconstruction quality achieved by a dome arrangement can be approximated with a fraction of light sources, if they are placed in a circular arrangement at an optimal elevation angle. The same is true for the theoretical and empirical analysis of errors in PS introduced by a simplified lighting model. While error estimation depending on size of the imaged object and light source distances is valid for arbitrary objects, the error mitigation strategy grounded in these findings is especially useful for recording graphical heritage applications: it is most effective for quasi-flat objects and most relevant if lighting distance cannot be adjusted arbitrarily, such as with a portable setup dome in a library/museum. For the considerations of error analysis and mitigation, only the Lambertian illumination model was considered. From an engineering-centered perspective on computer science (see Section 1.3), one could argue that recent learning-based PS methods [CHS+19, GMS+21, LSJ22, Ike23b, LZS+23] render such contributions obsolete: challenges in PS are handled satisfactorily without the need of explicitly modeling them. However, this is not the only valid view - in scientific traditions of computer science, there is intrinsic value in adding to the understanding of phenomena and processes. Furthermore, value in analytical models is found in their general applicability without domain-specific re-training and, in this particular case, in computational efficiency. In the case study presented in Section 4.4, the works on PS described above are applied and an effective visualization strategy for graphical heritage manifesting in minor depth variations is demonstrated.

### Developing objective evaluation methods for the quality of images visualizing graphical heritage

While one could argue that the usefulness of an image for accessing graphical heritage is only shown if it has been successfully used for that purpose (like in the case study described in Section 4.4), such specific observations do not allow generalizations about the performance of the imaging and image processing method used to create that image. Only a meta-study on a large number of case studies conducted using the same methods may justify claims of this kind - a very unpractical approach when trying to develop and evaluate new imaging/processing methods. The work presented in Chapter 5, on the other hand, is dedicated to finding methods for directly estimating the quality of

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

graphical heritage imagery, without the need for contextual information. In order to limit the scope, the quality criterion is reduced to text legibility.

Evaluating a dataset of artificially degraded manuscripts (Section 5.1) is a first step in this direction. While the MI with images before degradation *appears* to be a good proxy for perceived quality, a rigorous perceptual study is missing. However, as the applicability of such a reference comparison approach is limited to images of surfaces represented in expensively created datasets, it is not further investigated.

Instead, a novel dataset for the development of no-reference image assessment is introduced. The Subjective Assessments of Legibility in Ancient Manuscript Images (SALAMI) dataset contains 250 images of 50 manuscript patches, where in each image, the text regions are selected and rated for legibility by 20 scholars; these ratings are then averaged to obtain a legibility map for each image. Being created with a novel study design, the coherence and validity of the mean ratings obtained are statistically evaluated: reliable levels of both within- and between-rater variability are documented and the majority of variation is shown to originate from the surfaces being rated, rather than being attributed to systematic or random effects. This makes not only the dataset but also the study design used for its creation a contribution of its own. However, feedback from the participants revealed aspects of practical text deciphering that are not considered in the study design: the influence of line height on legibility, the time and effort for reading as an additional rating parameter, and the fact that scholars tend to dynamically adjust scale, contrast and brightness levels for deciphering text, instead of viewing a static image.

Based on the SALAMI dataset with its ground truth legibility scores, a framework for testing potential quantitative estimators for text legibility in images is tested. In this way, the use of the dataset is shown for evaluating whole-image estimations, spatially resolved estimations (that create legibility maps) and the performance of estimators in ranking different images of the same contents. None of the candidate methods tested shows a Spearman correlation with ground truth ratings greater than $\approx 0.7$ (which was reached by a dedicated convolutional neural network). Thus, no reliable quality estimator is found yet, but a baseline for further research is established.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

# Bibliography

[ACC+19] G. Adinolfi, R. Carmagnola, M. Cataldi, L. Marras, and V. Palleschi. Recovery of a Lost Wall Painting at the Etruscan Tomb of the Blue Demons in Tarquinia (Viterbo, Italy) by Multispectral Reflectometry and UV Fluorescence Imaging. *Archaeometry*, 61(2):450–458, April 2019.

[ACD18] C. T. C. Arsene, S. Church, and M. Dickinson. High Performance Software in Multidimensional Reduction Methods for Image Processing with Application to Ancient Manuscripts. *Manuscript Cultures*, 11:73–96, 2018.

[AFG13] J. Ackermann, S. Fuhrmann, and M. Goesele. Geometric Point Light Source Calibration. In *Vision, Modeling & Visualization*. The Eurographics Association, 2013.

[AG15] J. Ackermann and M. Goesele. A Survey of Photometric Stereo Techniques. *Foundations and Trends in Computer Graphics and Vision*, 9(3-4):149–254, 2015.

[AKMS12] T. O. Aydın, K. I. Kim, K. Myszkowski, and H. Seidel. NoRM : No-Reference Image Quality Metric for Realistic Image Synthesis. *Computer Graphics Forum*, 31(2), 2012.

[AMK07] N. G. Alldrin, S. P. Mallick, and D. J. Kriegman. Resolving the Generalized Bas-Relief Ambiguity by Entropy Minimization. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[APY16] S. A. Amirshahi, M. Pedersen, and S. X. Yu. Image Quality Assessment by Comparing CNN Features between Images. *Journal of Imaging Science and Technology*, 60(6):60410–1–60410–10, 2016.

[Arn14] D. Arnold. Computer Graphics and Cultural Heritage: From One-Way Inspiration to Symbiosis, Part 1. *IEEE Computer Graphics and Applications*, 34(3):76–86, May 2014.

[ASSS14] J. Ahmad, J. Sun, L. Smith, and M. Smith. An Improved Photometric Stereo through Distance Estimation and Light Vector Optimization from Diffused Maxima Region. *Pattern Recognition Letters*, 50:15–22, December 2014.

[Aud65] W. H. Auden. An Epithalamium for Peter Mudford & Rita Auden (May 15th, 1965). *The New Yorker*, page 34, July 1965.

[Aud69] W. H. Auden. Epistle to a Godson. *The New York Review of Books*, June 1969.

[Aud72] W. H. Auden. *Epistle to a Godson, And Other Poems*. Random House, New York, 1st ed. edition, 1972.

[AWL13] M. Aittala, T. Weyrich, and J. Lehtinen. Practical SVBRDF capture in the frequency domain. *ACM Transactions on Graphics*, 32(4), 2013.

[Bai90] P. Bailey. *General Certificate of Secondary Education Typewriting*. Macmillan Publishers Limited, 1990.

[Bar22] J. Barrett. ARCHiOx: Research and development in imaging. In *The Conveyor*. Bodleian Libraries, University of Oxford, May 2022.

[BCLP98] S. Baronti, A. Casini, F. Lotti, and S. Porcinai. Multispectral Imaging System for the Mapping of Pigments in Works of Art by Use of Principal-Component Analysis. *Applied Optics*, 37(8):1299–1309, March 1998.

[BCNS18] S. Bianco, L. Celona, P. Napoletano, and R. Schettini. On the use of deep learning for blind image quality assessment. *Signal, Image and Video Processing*, 12(2):355–362, 2018.

[BFM23] S. Brenner, T. Frühwirth, and S. Mayer. Revealing 'invisible' poetry by W. H. Auden through computer vision: Using photometric stereo to visualize indented impressions. *Digital Scholarship in the Humanities*, fqad037, June 2023.

[BKJ+17] S. S. Bukhari, A. Kadi, M. A. Jouneh, F. M. Mir, and A. Dengel. anyOCR: An Open-Source OCR System for Historical Archives. In *IAPR International Conference on Document Analysis and Recognition*, volume 01, pages 305–310, November 2017.

[BKY99] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The Bas-Relief Ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999.

[BMBW15] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[BMM+18] S. Bosse, D. Maniry, K.-R. Muller, T. Wiegand, and W. Samek. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, January 2018.

TU WIEN Bibliothek Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[Bre19] S. Brenner. On the Use of Artificially Degraded Manuscripts for Quality Assessment of Readability Enhancement Methods - Dataset & Code. https://doi.org/10.5281/zenodo.2650152, 2019.

[Bre20] S. Brenner. SALAMI - Subjective Assessments of Legibility in Ancient Manuscript Images. https://doi.org/10.5281/zenodo.4270352, 2020.

[Bre23] S. Brenner. Visualization of typewriter impressions in two letters of W. H. Auden. https://doi.org/10.5281/zenodo.7706092, 2023.

[BS21] S. Brenner and R. Sablatnig. Subjective Assessments of Legibility in Ancient Manuscript Images - The SALAMI Dataset. In *Pattern Recognition. ICPR International Workshops and Challenges*, volume 12667 of *Lecture Notes in Computer Science*, pages 68–82, Cham, 2021. Springer International Publishing.

[BSA08] J. Brauers, N. Schulte, and T. Aach. Multispectral Filter-Wheel Cameras: Geometric Distortion Model and Compensation Algorithms. *IEEE Transactions on Image Processing*, 17(12):2368–2380, December 2008.

[BZS18] S. Brenner, S. Zambanini, and R. Sablatnig. An Investigation of Optimal Light Source Setups for Photometric Stereo Reconstruction of Historical Coins. In *EUROGRAPHICS Workshop on Graphics and Cultural Heritage*, Vienna, 2018.

[CB13] M. Cavalieri and G. Baldini. *Oltre il riflesso. Storia, iconografia e società negli specchi etruschi del Museo Archeologico Nazionale di Parma.* Brepols Publishers, 2013.

[CBB03] D. M. Chabries, S. W. Booras, and G. H. Bearman. Imaging the Past: Recent Applications of Multispectral Imaging Technology to Deciphering Manuscripts. *Antiquity*, 77(296):359–372, June 2003.

[CCC+08] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. MeshLab: an Open-Source Mesh Processing Tool. In *Eurographics Italian Chapter Conference.* The Eurographics Association, 2008.

[CDL15] D. M. Conover, J. K. Delaney, and M. H. Loew. Automatic Registration and Mosaicking of Technical Images of Old Master Paintings. *Applied Physics A*, 119(4):1567–1575, June 2015.

[CE07] P. Campisi and K. Egiazarian. *Blind image deconvolution. Theory and applications.* CRC Press, Taylor & Francis Group, 2007.

[CG07] R. Clemens and T. Graham. *Introduction to Manuscript Studies.* Cornell university press, Itaca (N. Y.) London, 2007.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[CHS+19] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong. Self-Calibrating Deep Photometric Stereo Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8731–8739, Long Beach, CA, USA, June 2019. IEEE.

[Cla92] J. J. Clark. Active photometric stereo. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 29–34, June 1992.

[CMDPD23] D. Cimino, G. Marchioro, P. De Paolis, and C. Daffara. Evaluating the integration of Thermal Quasi-Reflectography in manuscript imaging diagnostic protocols to improve non-invasive materials investigation. *Journal of Cultural Heritage*, 62:72–77, July 2023.

[CMJ+22] I. Christova-Šomova, H. Miklas, D. Jordanov, S. Brenner, F. Cappa, B. Frühmann, W. Vetter, and M. Schreiner. *Apostolus Eninensis: Bibliothecae Nationalis Bulgaricae Codex 1144: Editio Nova*. Universitetsko izdatelstvo "Sv. Kliment Okhridski" ; Holzhausen Der Verlag, Sofii͡a : Wien, 2022.

[Cos15] A. Cosentino. Panoramic, Macro and Micro Multispectral Imaging: An Affordable System for Mapping Pigments on Artworks. *Journal of Conservation and Museum Studies*, 13(1):1–17, 2015.

[CPB+22] F. Cappa, G. Piñar, S. Brenner, B. Frühmann, W. Wetter, M. Schreiner, P. Engel, H. Miklas, and K. Sterflinger. The Kiev Folia: An interdisciplinary approach to unravelling the past of an ancient Slavonic manuscript. *International Biodeterioration & Biodegradation*, 167, February 2022.

[CRZ+22] Q. Chen, Y. Ren, Z. Zhao, W. Tao, and H. Zhao. Error Analysis of Photometric Stereo with Near Quasi-Point Lights. *Computer Graphics Forum*, 41(6):149–165, 2022.

[CSHB18] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 839–847, March 2018.

[Dav98] M. Davis. *Thinking Like an Engineer*. Oxford University Press, 1998.

[Dav12] E. R. Davies. *Computer and Machine Vision: Theory, Algorithms, Practicalities*. Elsevier, Amsterdam ; Boston, 4th ed edition, 2012.

[DB79] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, 1(2):224–227, February 1979.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[DC05] O. Drbohlav and M. Chantler. On Optimal Light Configurations in Photometric Stereo. In *IEEE International Conference on Computer Vision*, volume 2, pages 1707–1712, October 2005.

[DG85] N. T. De Grummond. The Etruscan Mirror. *Source: Notes in the History of Art*, 4(2/3):26–35, January 1985.

[DKS13] M. Diem, F. Kleber, and R. Sablatnig. Text Line Detection for Heterogeneous Documents. In *International Conference on Document Analysis and Recognition*, pages 743–747, August 2013.

[DMR+15] R. Dessì, C. Mannu, G. Rodriguez, G. Tanda, and M. Vanzi. Recent improvements in photometric stereo for rock art 3D imaging. *Digital Applications in Archaeology and Cultural Heritage*, 2(2):132–139, January 2015.

[DNT+09] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi. Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel. *International Workshop on Quality of Multimedia Experience*, pages 204–209, 2009.

[DS08] M. Diem and R. Sablatnig. Registration Of Ancient Manuscript Images Using Local Descriptors. *Digital Heritage, Proceedings of the 14th International Conferece on Virtual Systems and Multimedia*, pages 188–192, 2008.

[DSR+17] S. Dey, P. Shivakumara, K. S. Raghunandan, U. Pal, T. Lu, G. H. Kumar, and C. S. Chan. Script independent approach for multi-oriented text detection in scene image. *Neurocomputing*, 242:96–112, June 2017.

[Dun73] J. C. Dunn. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3):32–57, January 1973.

[DVC13] J. Dyer, G. Verri, and J. Cupitt. Multispectral Imaging in Reflectance and Photo-induced Luminescence modes: A User Manual. Technical report, 2013.

[DY13] R. Dosselmann and X. D. Yang. Improved method of finding the illuminant direction of a sphere. *Journal of Electronic Imaging*, 22(1), March 2013.

[ECK11] R. L. Easton, W. A. Christens-Barry, and K. T. Knox. Spectral image processing and analysis of the Archimedes Palimpsest. In *European Signal Processing Conference*, pages 1440–1444, August 2011.

[Ede07] A. H. Eden. Three Paradigms of Computer Science. *Minds and Machines*, 17(2):135–167, August 2007.

[EKC03] R. L. Easton, K. T. Knox, and W. A. Christens-Barry. Multispectral imaging of the Archimedes palimpsest. In *Applied Imagery Pattern Recognition Workshop*, pages 111–116, October 2003.

[EKC$^+$10] R. L. Easton, K. T. Knox, W. A. Christens-Barry, K. Boydston, M. B. Toth, D. Emery, and W. Noel. Standardized system for multispectral imaging of palimpsests. In *IS&T/SPIE Electronic Imaging*, San Jose, California, February 2010.

[EKCB18] R. L. Easton, K. T. Knox, W. A. Christens-Barry, and K. Boydston. Spectral Imaging Methods Applied to the Syriac Galen Palimpsest. *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies*, 3(1):69–82, 2018.

[End19] B. Endres. *Digitizing Medieval Manuscripts: The St. Chad Gospels, Materiality, Recoveries, and Representation in 2D & 3D*. Amsterdam University Press, 2019.

[FBF23] C. Falluomini, S. Brenner, and B. Frühmann. New Light on the Gothic Palimpsest from Bologna. In *Veröffentlichungen zur Byzanzforschung, Vienna*, pages 373–383, Vienna, 2023. Austrian Academy of Sciences Press.

[Fey75] P. Feyerabend. *Against Method: Outline of an Anarchistic Theory of Knowledge*. Verso Books, 1975.

[FGSS$^+$12] S. Faigenbaum-Golovin, B. Sober, A. Shaus, E. Piasetzky, G. Bearman, M. Cordonsky, and I. Finkelstein. Multispectral images of ostraca: Acquisition and analysis. *Journal of Archaeological Science*, 39:3581–3590, December 2012.

[Fia05] M. Fiala. ARTag, a fiducial marker system using digital techniques. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, II:590–596, 2005.

[FK06] C. Fischer and I. Kakoulli. Multispectral and hyperspectral imaging technologies in conservation: Current research and potential applications. *Studies in Conservation*, 51(sup1):3–16, June 2006.

[FQW$^+$17] H. Fan, L. Qi, N. Wang, J. Dong, Y. Chen, and H. Yu. Deviation correction method for close-range photometric stereo with nonuniform illumination. *Optical Engineering*, 56(10), October 2017.

[FSS$^+$12] S. Faigenbaum, B. Sober, A. Shaus, M. Moinester, E. Piasetzky, G. Bearman, M. Cordonsky, and I. Finkelstein. Multispectral images of ostraca: Acquisition and analysis. *Journal of Archaeological Science*, 39(12):3581–3590, 2012.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[FSSM05] A. R. Farooq, M. L. Smith, L. N. Smith, and S. Midha. Dynamic photometric stereo for on line quality control of ceramic tiles. *Computers in Industry*, 56(8-9):918–934, December 2005.

[GB16] D. Ghadiyaram and A. C. Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2016.

[GC16] R. Garg and S. Chaudhury. Automatic Selection of Parameters for Document Image Enhancement Using Image Quality Assessment. In *IAPR Workshop on Document Analysis Systems*, pages 422–427, April 2016.

[GCL+19] E. Grifoni, B. Campanella, S. Legnaioli, G. Lorenzetti, L. Marras, S. Pagnotta, V. Palleschi, F. Poggialini, E. Salerno, and A. Tonazzini. A new infrared true-color approach for visible-infrared multispectral image analysis. *Journal on Computing and Cultural Heritage*, 12(2), 2019.

[GCM+15] A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, and M. Terras. UCL Multispectral Processed Images of Parchment Damage Dataset. https://doi.org/10.14324/000.ds.1469099, 2015.

[GCM+17] A. Giacometti, A. Campagnolo, L. MacDonald, S. Mahony, S. Robson, T. Weyrich, M. Terras, and A. Gibson. The value of critical destruction: Evaluating multispectral image processing methods for the analysis of primary historical texts. *Digital Scholarship in the Humanities*, 32(1):101–122, 2017.

[GK16] D. Ghosh and N. Kaabouch. A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34:1–11, January 2016.

[GLD+18] T. Grüning, R. Labahn, M. Diem, F. Kleber, and S. Fiel. READ-BAD: A New Dataset and Evaluation Scheme for Baseline Detection in Archival Documents. In *IAPR International Workshop on Document Analysis Systems*, pages 351–356, April 2018.

[GLS+19] T. Grüning, G. Leifert, T. Strauß, J. Michael, and R. Labahn. A two-stage method for text line detection in historical documents. *International Journal on Document Analysis and Recognition*, 22(3):285–302, September 2019.

[GMLS11] M. Gau, H. Miklas, M. Lettner, and R. Sablatnig. Image Acquisition & Processing Routines for Damaged Manuscripts. *Digital Medievalist*, 6(0), March 2011.

[GMS+21] H. Guo, Z. Mo, B. Shi, F. Lu, S. K. Yeung, P. Tan, and Y. Matsushita. Patch-based Uncalibrated Photometric Stereo under Natural Illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[GNP09] B. Gatos, K. Ntirogiannis, and I. Pratikakis. ICDAR 2009 Document Image Binarization Contest (DIBCO 2009). In *International Conference on Document Analysis and Recognition*, pages 1375–1382, July 2009.

[HBS19] F. Hollaus, S. Brenner, and R. Sablatnig. CNN Based Binarization of MultiSpectral Document Images. In *International Conference on Document Analysis and Recognition*, pages 533–538, Sydney, Australia, September 2019. IEEE.

[HDF+15] F. Hollaus, M. Diem, S. Fiel, F. Kleber, and R. Sablatnig. Investigation of Ancient Manuscripts based on Multispectral Imaging. *Proceedings of the 2015 ACM Symposium on Document Engineering*, (1):93–96, 2015.

[HDS14] F. Hollaus, M. Diem, and R. Sablatnig. Improving OCR accuracy by applying enhancement techniques on multispectral images. *International Conference on Pattern Recognition*, pages 3080–3085, 2014.

[HGS12] F. Hollaus, M. Gau, and R. Sablatnig. Multispectral Image Acquisition of Ancient Manuscripts. In *Progress in Cultural Heritage Preservation*, Lecture Notes in Computer Science, pages 30–39, Berlin, Heidelberg, 2012. Springer.

[HJB+12] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, S. M. Brady, and J. A. Schnabel. MIND: Modality Independent Neighbourhood Descriptor for Multi-Modal Deformable Registration. *Medical Image Analysis*, 16(7):1423–1435, October 2012.

[HK04] I. Horovitz and N. Kiryati. Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing*, 22(9):681–694, August 2004.

[HM63] M. Herzberger and N. R. McClure. The Design of Superachromatic Lenses. *Applied Optics*, 2(6):553–560, 1963.

[HMV18] M. Hess, L. W. MacDonald, and J. Valach. Application of multi-modal 2D and 3D imaging and analytical techniques to document and examine coins on the example of two Roman silver denarii. *Heritage Science*, 6(1), December 2018.

[HNM+15] R. Hedjam, H. Z. Nafchi, R. F. Moghaddam, M. Kalacska, and M. Cheriet. ICDAR 2015 contest on MultiSpectral Text Extraction (MS-TEx 2015). In *International Conference on Document Analysis and Recognition*, pages 1181–1185, Tunis, Tunisia, August 2015. IEEE.

[Hol21] F. Hollaus. *Restoration of Multispectral Images of Ancient Documents*. PhD thesis, TU Wien, 2021.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[HRH+13] F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics*, 32(5):1–13, 2013.

[HS97] J. Heikkila and O. Silven. A Four-Step Camera Calibration Procedure with Implicit Image Correction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1106–1112, San Juan, Puerto Rico, 1997. IEEE Comput. Soc.

[HS17] F. Hollaus and R. Sablatnig. MultiSpectral Imaging for the Analysis of Historical Handwritings and Forgery Detection. In *Die getäuschte Wissenschaft*, pages 233–246. V&R unipress, Göttingen, 1 edition, May 2017.

[HW11] S. Herbort and C. Wöhler. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(3), September 2011.

[HZRS16] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, Las Vegas, NV, USA, June 2016. IEEE.

[Ike23a] S. Ikehata. Official pytorch repository for Scalable, Detailed and Mask-free Universal Photometric Stereo (CVPR2023). https://github.com/satoshi-ikehata/SDM-UniPS-CVPR2023, July 2023.

[Ike23b] S. Ikehata. Scalable, Detailed and Mask-Free Universal Photometric Stereo. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13198–13207, 2023.

[Int08] International Telecommunication Union. Subjective video quality assessment methods for multimedia applications P.910. *ITU-T*, April 2008.

[Int12] International Telecommunication Union. Methodology for the subjective assessment of the quality of television pictures ITU-R BT.500-13. *ITU-R*, January 2012.

[ISI90] Y. Iwahori, H. Sugie, and N. Ishii. Reconstructing shape from shading images under point light source illumination. In *International Conference on Pattern Recognition*, volume i, pages 83–87 vol.1, June 1990.

[JB91] X. Y. Jiang and H. Bunke. On error analysis for surface normals determined by photometric stereo. *Signal Processing*, 23(3):221–226, June 1991.

[JCT+19] C. Jones, W. A. Christens-Barry, M. Terras, M. B. Toth, and A. Gibson. Affine registration of multispectral images of historical documents for optimized feature recovery. *Digital Scholarship in the Humanities*, 35:587–600, July 2019.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[JDGT20] C. Jones, C. Duffy, A. Gibson, and M. Terras. Understanding multispectral imaging of cultural heritage: Determining best practice in MSI analysis of historical artefacts. *Journal of Cultural Heritage*, 45:339–350, September 2020.

[JF06] M. K. Johnson and H. Farid. Exposing digital forgeries through chromatic aberration. *Proceedings of the Multimedia and Security Workshop*, 2006(2):48–55, 2006.

[Joh19] M. Johnson. *Archaeological Theory: An Introduction*. John Wiley & Sons, March 2019.

[JYP07] M. Jackson, D. Yang, and R. Parkin. Analysis of wood surface waviness with a two-image photometric stereo method. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 221(8):1091–1099, December 2007.

[KCHM17] P. Kahle, S. Colutto, G. Hackl, and G. Mühlberger. Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents. In *IAPR International Conference on Document Analysis and Recognition*, volume 04, pages 19–24, November 2017.

[KCM21] T. Kleynhans, M. L. Carr, and D. W. Messinger. Low-cost, user friendly multispectral imaging system for the recovery of damaged, faded or palimpsested historical documents. In D. W. Messinger and M. Velez-Reyes, editors, *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVII*, page 8, Online Only, United States, April 2021. SPIE.

[KDL$^+$08] F. Kleber, M. Diem, M. Lettner, M. C. Vill, and R. Sablatnig. The Sinaitic Glagolitic Sacramentary Fragments. In *Electronic Imaging and the Visual Arts*, pages 23–29, Berlin, 2008.

[KDM$^+$18] E. C. Köhler, D. Driaux, S. Marchand, T. Holm, and A. Capirci. Preliminary Report on the Investigation of a Late Period Tomb with Aramaic Inscription at El-Sheikh Fadl/Egypt. *Egypt and the Levant*, 28:55–84, 2018.

[KE15] E. Kotoula and G. Earl. Digital Research Strategies for Ancient Papyri - A Case Study on Maounted Fragments of The Derveni Papyrus. In *CAA2014: 21st Century Archaeology: Concepts, Methods and Tools. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology*, pages 145–154. Archaeopress Publishing Ltd, March 2015.

[KGBN$^+$15] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V. R. Chandrasekhar, S. Lu, F. Shafait, S. Uchida, and E. Valveny. ICDAR 2015 competition on Robust Reading.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

In *International Conference on Document Analysis and Recognition*, pages 1156–1160, August 2015.

[KJW13] D. Kang, Y. J. Jang, and S. Won. Development of an inspection system for planar steel surface using multispectral photometric stereo. *Optical Engineering*, 52(3), March 2013.

[KL16] T. K. Koo and M. Y. Li. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2):155–163, 2016.

[KMB22] D. Kasotakis, P. Michael, and K. Boydston. Illuminating techniques from the sinai desert: Spectral imaging inside one of the oldest libraries in the world brings back to life re-written and forgotten texts. review of the imaging techniques and the global advantages on the field that were gained from the sinai palimpsests project. *Manuscript Cultures*, 15:87–90, 2022.

[Kno18] K. T. Knox. Image Processing Software for the Recovery of Erased or Damaged Text. *Manuscript Cultures*, 11:63–72, 2018.

[Kög14] P. R. Kögel. Die Palimpsestphotographie. Ein Beitrag zu den philologisch-historischen Hilfswissenschaften. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 37:974–978, 1914.

[KOMS11] M. Kobayashi, T. Okabe, Y. Matsushita, and Y. Sato. Surface Reconstruction in Photometric Stereo with Calibration Error. In *Visualization and Transmission 2011 International Conference on 3D Imaging, Modeling, Processing*, pages 25–32, May 2011.

[KSM$^+$10] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim. elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, Jan 2010.

[KSU$^+$13] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i. Bigorda, S. R. Mestre, J. Mas, D. F. Mota, J. A. Almazàn, and L. P. d. l. Heras. ICDAR 2013 Robust Reading Competition. In *International Conference on Document Analysis and Recognition*, pages 1484–1493, August 2013.

[Kuh62] T. S. Kuhn. *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press, 1962.

[KYLD14] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, 2014.

[LBMC20] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla. A CNN Based Approach for the Near-Field Photometric Stereo Problem. *arXiv:2009.05792*, September 2020.

[LBT+17] J. Liao, B. Buchholz, J.-M. Thiery, P. Bauszat, and E. Eisemann. Indoor Scene Reconstruction Using Near-Light Photometric Stereo. *IEEE Transactions on Image Processing*, 26(3):1089–1101, March 2017.

[LD19] T. Lu and A. Dooms. A Deep Transfer Learning Approach to Document Image Quality Assessment. In *International Conference on Document Analysis and Recognition*, pages 1372–1377, September 2019.

[LDSM08] M. Lettner, M. Diem, R. Sablatnig, and H. Miklas. Registration and enhancing of multispectral manuscript images. In *2008 16th European Signal Processing Conference*, 2008.

[LHS19] H. Lin, V. Hosu, and D. Saupe. KADID-10k: A Large-scale Artificially Distorted IQA Database. *International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3, 2019.

[LLS20] G. Leifert, R. Labahn, and J. A. Sánchez. Two Semi-Supervised Training Approaches for Automated Text Recognition. In *International Conference on Frontiers in Handwriting Recognition*, pages 145–150, September 2020.

[LND18] C. Liu, S. G. Narasimhan, and A. W. Dubrawski. Near-light photometric stereo using circularly placed point light sources. In *IEEE International Conference on Computational Photography*, pages 1–10, Pittsburgh, PA, May 2018. IEEE.

[LSB+17] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu. TextBoxes: A Fast Text Detector with a Single Deep Neural Network. In *AAAI Conference on Artificial Intelligence*, February 2017.

[LSDS11] L. Likforman-Sulem, J. Darbon, and E. H. Smith. Enhancement of historical printed document images by combining Total Variation regularization and Non-local Means filtering. *Image and Vision Computing*, 29(5):351–363, 2011.

[LSJ22] D. Lichy, S. Sengupta, and D. W. Jacobs. Fast Light-Weight Near-Field Photometric Stereo. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12612–12621, 2022.

[LTBB10] Z. Lu, Y.-W. Tai, M. Ben-Ezra, and M. S. Brown. A framework for ultra high resolution 3D imaging. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1205–1212, June 2010.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[LVDWB17] X. Liu, J. Van De Weijer, and A. D. Bagdanov. RankIQA: Learning from Rankings for No-Reference Image Quality Assessment. In *IEEE International Conference on Computer Vision*, pages 1040–1049, Venice, October 2017. IEEE.

[LW18] K.-Y. Lin and G. Wang. Hallucinated-IQA: No-Reference Image Quality Assessment via Adversarial Learning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 732–741, 2018.

[Lyo21] M. Lyons. *The Typewriter Century: A Cultural History of Writing Practices*. Number 46 in Studies in Book and Print Culture. University of Toronto press, Toronto Buffalo London, 2021.

[LZQ18] H. Li, F. Zhu, and J. Qiu. CG-DIQA: No-Reference Document Image Quality Assessment Based on Character Gradient. In *International Conference on Pattern Recognition*, pages 3622–3626, August 2018.

[LZQ19] H. Li, F. Zhu, and J. Qiu. Towards Document Image Quality Assessment: A Text Line Based Framework and a Synthetic Text Line Image Dataset. In *International Conference on Document Analysis and Recognition*, pages 551–558, September 2019.

[LZS+23] Z. Li, Q. Zheng, B. Shi, G. Pan, and X. Jiang. DANI-Net: Uncalibrated Photometric Stereo by Differentiable Shadow Handling, Anisotropic Reflectance Modeling, and Neural Inverse Rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8381–8391, 2023.

[MB11] A. K. Moorthy and A. C. Bovik. Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, 2011.

[MC03] G. McGunnigle and M. Chantler. Resolving handwriting from background printing using photometric stereo. *Pattern Recognition*, 36(8):1869–1879, August 2003.

[MCR18] H. Meißner, M. Cramer, and R. Reulke. Towards Standardized Evaluation of Image Quality for Airborne Camera Systems. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1:295–300, September 2018.

[MDA02] V. Masselus, P. Dutré, and F. Anrys. The free-form light stage. *Report CW* 335, Departement of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium, April 2002.

[Men22] E. Mendelson. Textual Notes. In *The Complete Works of W. H. Auden: Poems, Volume II: 1940-1973*, pages 782–1094. Princeton University Press, Princeton, 2022.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[MGC$^{+}$13] L. MacDonald, A. Giacometti, A. Campagnolo, S. Robson, T. Weyrich, M. Terras, and A. Gibson. Multispectral Imaging of Degraded Parchment. In *Computational Color Imaging*, volume 7786, pages 143–157. Springer, Berlin, Heidelberg, 2013.

[Min18] S. Mindermann. Hyperspectral Imaging for Readability Enhancement of Historic Manuscripts. Master's thesis, TU München, 2018.

[MJH19] A. Mathys, R. Jadinon, and P. Hallot. Exploiting 3D Multispectral Texture for a Better Feature Identification for Cultural Heritage. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W6:91–97, August 2019.

[MLBC21] R. Mecca, F. Logothetis, I. Budvytis, and R. Cipolla. LUCES: A Dataset for Near-Field Point Light Source Photometric Stereo. *arXiv:2104.13135*, October 2021.

[MMB12] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, December 2012.

[MN13] A. Miks and J. Novák. Method for primary design of superachromats. *Applied Optics*, 52(28):6868–6876, 2013.

[MS15] R. A. Manap and L. Shao. Non-distortion-specific no-reference image quality assessment: A survey. *Information Sciences*, 301:141–160, April 2015.

[MST$^{+}$19] G. Muehlberger, L. Seaward, M. Terras, S. Ares Oliveira, V. Bosch, M. Bryan, S. Colutto, H. Déjean, M. Diem, S. Fiel, B. Gatos, A. Greinoecker, T. Grüning, G. Hackl, V. Haukkovaara, G. Heyer, L. Hirvonen, T. Hodel, M. Jokinen, P. Kahle, M. Kallio, F. Kaplan, F. Kleber, R. Labahn, E. M. Lang, S. Laube, G. Leifert, G. Louloudis, R. McNicholl, J.-L. Meunier, J. Michael, E. Mühlbauer, N. Philipp, I. Pratikakis, J. Puigcerver Pérez, H. Putz, G. Retsinas, V. Romero, R. Sablatnig, J. A. Sánchez, P. Schofield, G. Sfikas, C. Sieber, N. Stamatopoulos, T. Strauß, T. Terbul, A. H. Toselli, B. Ulreich, M. Villegas, E. Vidal, J. Walcher, M. Weidemann, H. Wurster, and K. Zagoris. Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation*, 75(5):954–976, September 2019.

[MTM12] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk. Comparison of four subjective methods for image quality assessment. *Computer Graphics Forum*, 31(8):2478–2491, 2012.

[Mus95] S. Musulin. Auden in Kirchstetten. In *In Solitude, for Company: W.H. Auden after 1940, Unpublished Prose and Recent Criticism*, pages 207–233. Clarendon Press, Oxford, 1995.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[MWBK14] R. Mecca, A. Wetzler, A. M. Bruckstein, and R. Kimmel. Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences*, 7(4):2732–2770, January 2014.

[NIK90] S. K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431, 1990.

[NM12] L. Neumann and J. Matas. Real-time scene text localization and recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3538–3545, Providence, RI, June 2012. IEEE.

[Nol29] R. Noll. Ein neuer etruskischer Spiegel. *Mitteilungen des Vereines Klassischer Philologen in Wien*, VI. Jahrgang:39–47, 1929.

[NS16] Y. Nie and Z. Song. A novel photometric stereo method with nonisotropic point light sources. In *International Conference on Pattern Recognition*, pages 1737–1742, Cancun, December 2016. IEEE.

[NSJZ16] Y. Nie, Z. Song, M. Ji, and L. Zhu. A novel calibration method for the photometric stereo system with non-isotropic LED lamps. In *IEEE International Conference on Real-time Computing and Robotics*, pages 289–294, June 2016.

[OAA12] T. Obafemi-Ajayi and G. Agam. Character-Based Automated Human Perception Quality Assessment in Document Images. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 42(3):584–595, May 2012.

[PCCS20] M. Picollo, C. Cucci, A. Casini, and L. Stefani. Hyper-Spectral Imaging Technique in the Cultural Heritage Field: New Possible Scenarios. *Sensors*, 20(10), January 2020.

[PDG+17] E. Pouyet, S. Devine, T. Grafakos, R. Kieckhefer, J. Salvant, L. Smieska, A. Woll, A. Katsaggelos, O. Cossairt, and M. Walton. Revealing the biography of a hidden medieval manuscript using synchrotron and conventional imaging techniques. *Analytica Chimica Acta*, 982:20–30, 2017.

[PF13] T. Papadhimitri and P. Favaro. A New Perspective on Uncalibrated Photometric Stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1474–1481, Portland, OR, USA, June 2013. IEEE.

[PF14] T. Papadhimitri and P. Favaro. Uncalibrated Near-Light Photometric Stereo. In *British Machine Vision Conference*, pages 128.1–128.12, Nottingham, 2014. British Machine Vision Association.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[PF16] D. Perrone and P. Favaro. A Clearer Picture of Total Variation Blind Deconvolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(6):1041–1055, 2016.

[PG01] E. Pringsheim and O. Gradenwitz. Photographische Reconstruction von Palimpsesten. *Jahrbuch für Photographie und Reproduktionstechnik*, 15(1991):52–56, 1901.

[Piq17] K. E. Piquette. Illuminating the Herculaneum Papyri: Testing new imaging techniques on unrolled carbonised manuscript fragments. *Digital Classics Online*, pages 80–102, November 2017.

[PJI$^+$15] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C. C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.

[PLZ$^+$09] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. TID2008 - A database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45, 2009.

[PMd23] V. M. Papadakis, M. Machado, and J. dos Santos. XpeCAM: The Complete Solution for Artwork Documentation and Analysis. In *The Future of Heritage Science and Technologies*, Lecture Notes in Mechanical Engineering, pages 16–27, Cham, 2023. Springer International Publishing.

[PNR15] M.-T. Puth, M. Neuhäuser, and G. D. Ruxton. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, 102:77–84, April 2015.

[POMZ$^+$19] M. Perez-Ortiz, A. Mikhailiuk, E. Zerman, V. Hulusic, G. Valenzise, and R. K. Mantiuk. From pairwise comparisons and rating to a unified quality scale. *IEEE Transactions on Image Processing*, 29:1139–1151, 2019.

[PSG01] M. Powell, S. Sarkar, and D. Goldgof. A simple strategy for calibrating the geometry of light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):1022–1027, September 2001.

[PZK$^+$19] I. Pratikakis, K. Zagoris, X. Karagiannis, L. Tsochatzidis, T. Mondal, and I. Marthot-Santaniello. ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019). In *International Conference on Document Analysis and Recognition*, pages 1547–1556, September 2019.

[QDW$^+$18] Y. Quéau, B. Durix, T. Wu, D. Cremers, F. Lauze, and J.-D. Durou. LED-Based Photometric Stereo: Modeling, Calibration and Numerical Solution. *Journal of Mathematical Imaging and Vision*, 60(3):313–340, March 2018.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

[QLD15] Y. Quéau, F. Lauze, and J.-D. Durou. Solving Uncalibrated Photometric Stereo Using Total Variation. *Journal of Mathematical Imaging and Vision*, 52(1):87–107, May 2015.

[Qui13] J. Quinn. At Home in Italy and Austria, 1948–1973. In *W. H. Auden in Context*, pages 56–66. Cambridge University Press, January 2013.

[Qui15] J. Quinn. Auden's Cold War Fame. In *Auden at Work*, pages 231–249. Palgrave Macmillan, Basingstoke, 2015.

[R D08] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.

[RBC+19] C. Römer, S. Brenner, F. Cappa, B. Frühmann, E.-G. Hammerschmid, and M. Schreiner. Recovering a 16th-century Ottoman document damaged by spilled ink. In *International Conference El'Manuscript „Textual Heritage and Information Technologies"*. Gutenberg Publishing House, 2019.

[RBK83] R. Ray, J. Birk, and R. B. Kelley. Error Analysis of Surface Normals Determined by Radiometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(6):631–645, November 1983.

[RFN11] F. Ribeiro, D. Florencio, and V. Nascimento. Crowdsourcing subjective image quality evaluation. *International Conference on Image Processing*, pages 3097–3100, 2011.

[RKC14] J. Ryu, H. I. Koo, and N. I. Cho. Language-Independent Text-Line Extraction Algorithm for Handwritten Documents. *IEEE Signal Processing Letters*, 21(9):1115–1119, September 2014.

[RL15] M. Rosenberger and G. Linß. Multispectral Image Correction for Geometric Measurements. *Journal of Physics: Conference Series*, 588:012037, February 2015.

[SB06] H. R. Sheikh and A. C. Bovik. Image Information and Visual Quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[SBAG22] F. A. Saiz, I. Barandiaran, A. Arbelaiz, and M. Graña. Photometric Stereo-Based Defect Detection System for Steel Components Manufacturing Using a Deep Segmentation Network. *Sensors*, 22(3):882, January 2022.

[SBL+14] D. Shamonin, E. Bron, B. Lelieveldt, M. Smits, S. Klein, and M. Staring. Fast Parallel Image Registration on CPU and GPU for Diagnostic Classification of Alzheimer's Disease. *Frontiers in Neuroinformatics*, 7:50, 2014.

TU Bibliothek WIEN Your knowledge hub
Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[SC06] A. Spence and M. Chantler. Optimal illumination for three-image photometric stereo using sensitivity analysis. *IEE Proceedings - Vision, Image, and Signal Processing*, 153(2):149, 2006.

[Sch97] K. Schlüns. The Irradiance Error and its Effect in Photometric Stereo. *Communication and Information Technology Research Technical Report* 13, The University of Auckland, 1997. http://hdl.handle.net/2292/2758.

[SCS90] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):435–446, May 1990.

[SF79] P. E. Shrout and J. L. Fleiss. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420–428, 1979.

[SF11] M. Stommel and G. Frieder. Automatic Estimation of the Legibility of Binarised Historic Documents for Unsupervised Parameter Tuning. In *International Conference on Document Analysis and Recognition*, pages 104–108, September 2011.

[SFGST17] A. Shaus, S. Faigenbaum-Golovin, B. Sober, and E. Turkel. Potential Contrast – A New Image Quality Measure. *Electronic Imaging*, 2017(12):52–58, 2017.

[ŠKB+10] Ž. Špiclin, J. Katrašnik, M. Bürmen, F. Pernuš, and B. Likar. Geometric calibration of a hyperspectral imaging system. *Applied Optics*, 49(15):2813–2818, 2010.

[Sli16] D. H. Sliney. What is light? The visible spectrum and beyond. *Eye*, 30(2):222–229, February 2016.

[SNAMC18] A. Shahkolaei, H. Z. Nafchi, S. Al-Maadeed, and M. Cheriet. Subjective and objective quality assessment of degraded document images. *Journal of Cultural Heritage*, 30:199–209, March 2018.

[SON19] A. Sulaiman, K. Omar, and M. F. Nasrudin. Degraded Historical Document Binarization: A Review on Issues, Challenges, Techniques, and Future Directions. *Journal of Imaging*, 5(4):48, April 2019.

[SSB06] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, 15(11):3441–3452, 2006.

[SSS+07] J. Sun, M. Smith, L. Smith, S. Midha, and J. Bamber. Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities. *Image and Vision Computing*, 25(7):1050–1057, July 2007.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

[SSSF07] J. Sun, M. Smith, L. Smith, and A. Farooq. Examining the uncertainty of the recovered surface normal in three light photometric stereo. *Image and Vision Computing*, 25(7):1073–1079, July 2007.

[SSSF13] J. Sun, M. Smith, L. Smith, and A. Farooq. Sampling Light Field for Photometric Stereo. *International Journal of Computer Theory and Engineering*, pages 14–18, 2013.

[ST97] G. Sharma and H. Trussell. Digital color imaging. *IEEE Transactions on Image Processing*, 6(7):901–932, July 1997.

[STB07] E. Salerno, A. Tonazzini, and L. Bedini. Digital image analysis to enhance underwritten text in the Archimedes palimpsest. *International Journal on Document Analysis and Recognition*, 9(2-4):79–87, 2007.

[SWCB05] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik. LIVE Image Quality Assessment Database Release 2. http://live.ece.utexas.edu/research/quality, 2005.

[SWL+20] H. Santo, M. Waechter, W.-Y. Lin, Y. Sugano, and Y. Matsushita. Light Structure from Pin Motion: Geometric Point Light Source Calibration. *International Journal of Computer Vision*, 128(7):1889–1912, July 2020.

[SWM+16] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, and P. Tan. A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3707–3716, June 2016.

[The21] The MathWorks Inc. Computer Vision Toolbox version: 10.1 (R2021b), 2021.

[TPSE13] S. Thumfart, W. Palfinger, M. Stoger, and C. Eitzinger. Accurate Fibre Orientation Measurement for Carbon Fibre Surfaces. In *International Conference on Computer Analysis of Images and Patterns*, Lecture Notes in Computer Science, 2013.

[Tra13] J. Traa. Least-squares intersection of lines. *University of Illinois Urbana-Champaign (UIUC)*, 2013.

[TS08] M. Tedre and E. Sutinen. Three traditions of computing: What educators should know. *Computer Science Education*, 18(3):153–170, September 2008.

[TSA+19] A. Tonazzini, E. Salerno, Z. A. Abdel-Salam, M. A. Harith, L. Marras, A. Botto, B. Campanella, S. Legnaioli, S. Pagnotta, F. Poggialini, and V. Palleschi. Analytical and mathematical methods for revealing hidden details in ancient manuscripts and paintings: A review. *Journal of Advanced Research*, 17:31–42, May 2019.

[VBMR15] M. Vanzi, P. E. Bagnoli, C. Mannu, and G. Rodriguez. Photometric Stereo 3D Visualizations of Rock-Art Panels, Bas-Reliefs, and Graffiti. In *Conference on Computer Applications and Quantitative Methods in Archaeology*, 2015.

[VHW+18] B. Vandermeulen, H. Hameeuw, L. Watteeuw, L. Van Gool, and M. Proesmans. Bridging Multi-light & Multi-Spectral images to study, preserve and disseminate archival documents. *Archiving Conference*, 15(1):64–69, April 2018.

[VNV+15] T. Virtanen, M. Nuutinen, M. Vaahteranoksa, P. Oittinen, and J. Häkkinen. CID2013: A database for evaluating no-reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 24(1):390–402, 2015.

[VWI97] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, Sep 1997.

[WBS+04] Z. Wang, A. C. Bovik, H. R. Sheikh, S. Member, E. P. Simoncelli, and S. Member. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[WCCM00] G. Ware, D. Chabries, R. Christiansen, and C. Martin. Multispectral document enhancement: Ancient carbonized scrolls. In *International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment*, volume 6, pages 2486–2488, July 2000.

[Weg76] P. Wegner. Research paradigms in computer science. In *Proceedings of the 2nd International Conference on Software Engineering*, pages 322–330, Washington, DC, USA, October 1976. IEEE Computer Society Press.

[WGS+11] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust Photometric Stereo via Low-Rank Matrix Completion and Recovery. In *Asian Conference on Computer Vision*, pages 703–717. Springer, Berlin, Heidelberg, 2011.

[Woo80] R. J. Woodham. Photometric Method For Determining Surface Orientation From Multiple Images. *Optical Engineering*, 19(1), February 1980.

[WZSE15] E. Weigl, S. Zambal, M. Stöger, and C. Eitzinger. Photometric stereo sensor for robot-assisted industrial quality inspection of coated composite material surfaces. In *International Conference on Quality Control by Artificial Vision*, pages 367–374. SPIE, April 2015.

[XCW15] W. Xie, Chengkai Dai, and C. C. L. Wang. Photometric stereo with near point lighting: A solution by mesh deformation. In *IEEE Conference on*

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.
TU WIEN Bibliothek Your knowledge hub

*Computer Vision and Pattern Recognition*, pages 4585–4593, Boston, MA, USA, June 2015. IEEE.

[XLL20] X. Xu, L. Liu, and B. Li. A survey of CAPTCHA technologies to distinguish between human and computer. *Neurocomputing*, 408:292–307, September 2020.

[XNSW19] W. Xie, Y. Nie, Z. Song, and C. C. L. Wang. Mesh-Based Computation for Solving Photometric Stereo With Near Point Lighting. *IEEE Computer Graphics and Applications*, 39(3):73–85, May 2019.

[XSC13] W. Xie, Z. Song, and R. C. Chung. Real-time three-dimensional fingerprint acquisition via a new photometric stereo means. *Optical Engineering*, 52(10):103103, October 2013.

[YD13a] P. Ye and D. Doermann. Document Image Quality Assessment: A Brief Survey. In *International Conference on Document Analysis and Recognition*, pages 723–727, August 2013.

[YD13b] P. Ye and D. Doermann. Combining preference and absolute judgements in a crowd-sourced setting. *Proceedings of International Conference on Machine Learning*, pages 1–7, 2013.

[YKKD13] P. Ye, J. Kumar, L. Kang, and D. Doermann. Real-Time No-Reference Image Quality Assessment Based on Filter Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 987–994, June 2013.

[YMH$^+$16] C.-K. Yeh, N. Matsuda, X. Huang, F. Li, M. Walton, and O. Cossairt. A Streamlined Photometric Stereo Framework for Cultural Heritage. In *Computer Vision – ECCV 2016 Workshops*, volume 9913, pages 738–752. Springer International Publishing, Cham, 2016.

[YSB$^+$15] M. R. Yousefi, M. R. Soheili, T. M. Breuel, E. Kabir, and D. Stricker. Binarization-free OCR for historical documents using LSTM networks. In *International Conference on Document Analysis and Recognition*, pages 1121–1125, August 2015.

[ZMM10] X. Zhu, S. Member, and P. Milanfar. Automatic Parameter Selection for Denoising Algorithms Using a No-Reference Measure of Image Content. *IEEE Transactions on Image Processing*, 19(12):3116–3132, 2010.

[ZPE16] S. Zambal, W. Palfinger, and C. Eitzinger. Robotic inspection of 3D CFRP surfaces. In *IEEE Metrology for Aerospace*, pages 197–202, Florence, Italy, June 2016. IEEE.

[ZYW$^+$17] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang. EAST: An Efficient and Accurate Scene Text Detector. In *IEEE Conference on*

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

*Computer Vision and Pattern Recognition*, pages 2642–2651, Honolulu, HI, July 2017. IEEE.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

# Appendix

## Acronyms

Acronyms used in this thesis are listed below, with references to chapters/sections in which they occur.

| | | |
|---|---|---|
| **CLAHE** | Contrast Limited Adaptive Histogram Equalization | 5.1 |
| **CNN** | Convolutional Neural Network | 5.3 |
| **ICC** | Intraclass Correlation Coefficient | 5.2 |
| **IQA** | Image Quality Assessment | 2.3, 5 |
| **IR** | InfraRed | 2.1, 3.1, 3.2 |
| **MAE** | Mean Angular Error | 4.2, 4.3 |
| **MI** | Mutual Information | 5.1 |
| **MIND** | Modality Independent Neighborhood Descriptor | 3.2.4 |
| **MS** | MultiSpectral | 2.1, 3, 5.1, 5.2, 6 |
| **OCR** | Optical Character Recognition | 2.3, 5 |
| **PCC** | Pearson Correlation Coefficient | 5.1 |
| **PS** | Photometric Stereo | 2.2, 3, 4, 6 |
| **RMSE** | Root Mean Square Error | 3.2 |
| **SALAMI** | Subjective Assessments of Legibility in Ancient Manuscript Images | 5, 6 |
| **SMI** | Scientific Manuscript Images | 5.2 |
| **SRC** | Spearman Rank Correlation coefficient | 5.3 |
| **SSIM** | Structural SIMilarity index | 5.1 |
| **UV** | UltraViolet | 2.1, 3.1, 3.2 |
| **VIF** | Visual Information Fidelity | 5.1 |

Die approbierte gedruckte Originalversion dieser Dissertation ist an der TU Wien Bibliothek verfügbar.
The approved original version of this doctoral thesis is available in print at TU Wien Bibliothek.

## Publications

The following peer-reviewed publications were published in the course of my doctoral studies, in chronological order:

**Simon Brenner**, Sebastian Zambanini and Robert Sablatnig. An Investigation of Optimal Light Source Setups for Photometric Stereo Reconstruction of Historical Coins. *EUROGRAPHICS Workshop on Graphics and Cultural Heritage*, 2018.

**Simon Brenner** & Robert Sablatnig. On the Use of Artificially Degraded Manuscripts for Quality Assessment of Readability Enhancement Methods. *Proceedings of the ARW & OAGM Workshop*, 2019.

Fabian Hollaus, **Simon Brenner**, & Robert Sablatnig. CNN Based Binarization of MultiSpectral Document Images. *International Conference on Document Analysis and Recognition (ICDAR)*, 2019.

**Simon Brenner** & Robert Sablatnig. Lens Calibration for Focus Shift Correction in Close-Range Multispectral Imaging. *EUROGRAPHICS Workshop on Graphics and Cultural Heritage*, 2019.

**Simon Brenner** & Robert Sablatnig. Subjective Assessments of Legibility in Ancient Manuscript Images — The SALAMI Dataset. *Pattern Recognition. ICPR International Workshops and Challenges*, 2021.

**Simon Brenner**, Lukas Schügerl & Robert Sablatnig. Estimating Human Legibility in Historic Manuscript Images — A Baseline. *International Conference on Document Analysis and Recognition (ICDAR)*, 2021.

Federica Cappa, Guadalupe Piñar, **Simon Brenner**, Bernadette Frühmann, Wilfried Vetter, Manfred Schreiner, Patricia Engel, Heinz Miklas & Katja Sterflinger. The Kiev Folia: An interdisciplinary approach to unravelling the past of an ancient Slavonic manuscript. *International Biodeterioration & Biodegradation*, 167, 105342, 2022.

**Simon Brenner**, Timo Frühwirth, Sandra Mayer. Revealing 'invisible' poetry by W. H. Auden through computer vision: Using photometric stereo to visualize indented impressions. *Digital Scholarship in the Humanities*, fquad037, 2023.

**Simon Brenner** & Robert Sablatnig. Classical Photometric Stereo in Point Lighting Environments: Error Analysis and Mitigation. Accepted for *International Conference on 3D Vision*, 2024.