PhD thesis 2017

Cultural Heritage Acquisition: Obtaining Accurate Surface Geometry and Radiometry in the Wild

Thomas Höll

Supervisor: Ao.Univ.-Prof.Dr. Axel Pinz

Institute of Electrical Measurement
and Measurement Signal Processing
Graz University of Technology
Graz, Austria 2017

Statutory declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Graz, September 13, 2017
place, date

signature

I dedicate this thesis to my beloved and helpful wife Andrea and my joyful kids, Ben, Lena and Mona. They supported me in times of struggle and reminded me of what really matters in life.

O taste and see that the LORD is good:
blessed is the man that
trusteth in him.
Psalm 34,8 [KJV]

Acknowledgements

I express my profound thanks to Axel Pinz, who supported and supervised me with immense patience and understanding. Without his skilled advice and personal support this thesis would not have been possible.
I also wish to thank Matteo Dellepiane for revising my thesis in such a motivating way.
During the work within the 3D Pitoti project I learned to value some of my colleges better. I thank Gert Holler for his support, ideas and knowledge, especially while developing the scanner prototype. I also want to thank my colleagues from ICG for their support with the SfM software and the interesting insights in their work. Big thanks also to Martin Schaich, Oliver Reuß and the staff of ArcTron for their valuable feedback concerning the scanner.
I am more than thankful to all my colleagues and friends at the Institute. So many good ideas, helpful advice, cheerful moments, and motivating understanding resulted from the *uncountable* coffee breaks.
Last but not least my deep gratitude belongs to my family, my helpful wife and my joyful kids, who also supported me and for their understanding for my work late in the evening or on the weekend.

Cultural Heritage Acquisition:
Obtaining Accurate Surface Geometry and Radiometry in the Wild

Thomas Höll
Institute of Electrical Measurement and Measurement Signal Processing
Graz University of Technology

Abstract

Cultural heritage acquisition "in the wild" is challenging, because one is confronted with a non-controllable environment. In this thesis I tackle this challenge and present my approach to acquire accurate surface geometry and radiometry in three parts.
For geometry measurement, I extend Structure-from-Motion with constrained bundle adjustment postprocessing to obtain geo-referenced Euclidean reconstructions. As constraints for this task I use the geometric calibration of a newly developed scanner prototype and optionally available tachymeter measurements.
For radiometry measurement, I use the geometrically calibrated high-intensity illumination of the scanner and the previously obtained geometry. Calculating the relations between surface normal, camera pose, and incident light leads to the final result of dense 3D point clouds that represent radiometric surface properties.
For large sites, "true" radiometry for the complete 3D point cloud is very expensive to obtain. As my third contribution, I present a method that is capable to reconstruct the radiometric surface properties of an entire scene despite the fact that only for a part of it the "true" radiometry is known. This is done in a two stage process: First, I transfer the radiometry to spatially corresponding parts of the scene, and second, these values are propagated to the entire scene using affinity information.
The presented approach was used in the European 3D Pitoti project and showed its applicability for cultural heritage acquisition. Besides cultural heritage acquisition I validate the performance of my approach on various other objects and data.

Kurzfassung

Die bildbasierte Erfassung von Kulturerbe in ihrem natürlichen Umfeld ist eine anspruchsvolle Aufgabe. Ein Grund dafür ist die nicht kontrollierbare Umgebung. In dieser Dissertation behandle ich diese Problemstellung und zeige meine Methode auf, die es erlaubt die Geometrie und Radiometrie von Kulturgut akkurat zu erfassen.

Die Geometrie wird zuerst mit Hilfe von "Structure-from-Motion" (SfM) rekonstruiert. Im Anschluss daran werden mit Hilfe eines "constrained bundle ajustment" Verfahrens (Blockbündelausgleich unter Randbedingungen) etwaige Totalstation Messungen und die Kalibration eines neuentwickelten Scanners mit dem Ergebnis der SfM Rekonstruktion fusioniert.

Unter Einbeziehen der rekonstruierten Geometrie (inkl. Oberflächennormale und Kamera Posen) und der geometrisch kalibrierten Beleuchtung des Scanners wird die radiometrische Oberflächeneigenschaft rekonstruiert. Das Ergebnis dieser Berechnung ist eine dichte 3D Punktwolke, die die radiometrischen Oberflächeneigenschaften beinhaltet.

Für große Areale ist eine radiometrische Rekonstruktion sehr zeitaufwendig und teuer. Als drittes stelle ich eine Methode vor, die die Radiometrie eines gesamten Areals rekonstruiert, obwohl nur einzelne Ausschnitte radiometrisch erfasst wurden. Dazu wird zuerst die Radiometrie auf die korrespondierenden Bereiche der Szene übertragen, und im Anschluss auf den übrigen Teil weiterpropagiert.

Im Rahmen des EU-Projektes "3D Pitoti" wurde die präsentierte Methode zur bildbasierten Erfassung von Kulturerbe eingesetzt und zeigte ihre praktische Anwendbarkeit. Zusätzlich validierte ich die Methode unter Zuhilfenahme verschiedener Objekte und Datensätze.

Contents

List of Figures

List of Tables

List of Symbols

Geometric Symbols

$\boldsymbol{p}$	(m)	Surface point or point in $\mathbb{R}^3$
$\boldsymbol{\omega}$	(m)	Unit vector; used for directions
$\boldsymbol{n_p}$	(m)	Unit vector; Surface normal at point $\boldsymbol{p}$
$\boldsymbol{p'}$	$(Pixel)$	Corresponding pixel of point $\boldsymbol{p}$
$\boldsymbol{R}$	$(.)$	Rotation matrix $\boldsymbol{R} \in \mathfrak{so}(3)$
$\boldsymbol{t}$	(m)	Translation vector $\boldsymbol{t} \in \mathbb{R}^3$
$\boldsymbol{c}$	(m)	Camera center $\boldsymbol{c} \in \mathbb{R}^3$
$\boldsymbol{K}$	$(.)$	Camera calibration matrix $\boldsymbol{K} \in \mathbb{R}^{3\times3}$
$\boldsymbol{P}$	$(.)$	Camera projection matrix $\boldsymbol{P} \in \mathbb{R}^{3\times4}$
$\boldsymbol{m}$	(m)	Tachymeter measurement $\boldsymbol{m} \in \mathbb{R}^3$

Radiometric and Photometric Symbols

L	$\left(\frac{W}{sr \cdot m^2}\right)$	Radiance
I	$\left(\frac{W}{sr}\right)$	Radiant intensity
$f(\boldsymbol{p}, \boldsymbol{\omega_o}, \boldsymbol{\omega_i})$	$\left(\frac{1}{sr}\right)$	Bidirectional reflectance distribution functions (BRDF)
$\rho_c(\boldsymbol{p})$	$\left(\frac{1}{sr}\right)$	Lambertian reflectance property of a surface point $\boldsymbol{p}$ or its corresponding image pixel $\boldsymbol{p'}$ for the color channel c

$P_c(\boldsymbol{p})$ or $P_c(\boldsymbol{p'})$	(.)	Measured colour value of a surface point $\boldsymbol{p}$ or its corresponding image pixel $\boldsymbol{p'}$ for the color channel c

1

Introduction

Preserving and studying our past is a major field in science and leads to interesting multidisciplinary research questions and answers. Computer science aids archaeologists to answer questions about the past or enables quantitative methods to compare different hypotheses. Today we have the possibility to store a huge amount of data and have efficient algorithms to query it. With the advent of high-performance GPUs [1] and their capability of massive parallel processing we are able to capture, process and visualize large scale data such as point clouds or high resolution images. These and other innovations result in interesting and successful projects in the field of cultural heritage where archaeologists and computer scientists use their expertise to drive research in both areas.
This thesis arose in the context of such a setting. As part of a European project I was

[1] Graphical Processing Unit

concerned with the research topic of geometry and radiometry acquisition "in the wild" in the context of cultural heritage of rock-art sites.

1.1 Context of the thesis - The 3D Pitoti Project

This thesis arose in the context of the European project 3D Pitoti [2]. As a result, the projects aim and its requirements influence large parts of my work. The 3D Pitoti project aimed at scanning of large rock-art sites at various scales and the spatial relation between them, applied to the ancient rock-art site of Valcamonica, which contains at least $100,000$ petroglyphs ([14], [15], see figure 1.1 for some examples of different petroglyphs). The projects name is derived from the Camunian dialect word "Pitoti" which has the meaning "small puppets" and is the name of the figures [27] depicted by a huge number of petroglyphs (see figure 1.1).

The scanning itself is done at various scales to capture the ancient rock-art starting from individual petroglyphs (micro-scale) through to parts of the valley (a more general view of the multi-scale reconstruction approach can be found in [13], [103] and section 3.1 of this thesis).

Furthermore, the aim of the project is not only on 3D scanning that results in point clouds and meshes, but also on novel use cases of the 3D data in various applications including large-scale visualization [80], pecking style segmentation [138], petroglyph classification [174], and educational games [82] to name a few. In this context our part was the micro-scale scanning of individual petroglyphs in terms of their geometry (including their geo-referenced location) and radiometry. This task leads me to the following problem statement:

1.2 Problem statement

In its three main parts, my thesis addresses the following three research questions:

How can we produce accurate geo-referenced geometric reconstructions of the rock art?
The geometric reconstruction is a core part of the project and also of this thesis. In terms of accuracy, the reconstruction error is better than 0.1mm and the resolution of an individual piece of rock-art is about 0.04mm. To establish the geometric context of an individual piece of rock-art the proposed method also allows that the final reconstruction can be geo-referenced. Chapter 4 describes the proposed solution in detail.

Figure 1.1: Images of Pitoti (petroglyphs). "Pitoti" is a Camunian dialect word of which one meaning is "small puppets". As one can see there exists a variety of different shapes.

How can we capture its radiometry "in the wild"?

High-quality visualization requires that the radiometric surface properties are also captured without any influence of ambient illumination during scanning, *e.g.* cast shadows that will cause "double-shadow" artefacts during visualization. I want to emphasise the term "in the wild". The scanning is done outdoors and not in controlled laboratories. While this is a usual setting for the geometric reconstruction, it is a challenging task for the radiometric reconstruction because of the incident illumination from the surrounding

environment. Chapter 5 presents the developed method.

How can we propagate measured radiometry into areas where we have only access to photo texture?
Despite the fact that the proposed method to capture the radiometry works rather fast, collecting radiometric surface properties of an entire scene is often prohibitive. To provide consistent radiometry for a large scene, I developed a method to propagate known radiometric surface properties into regions were no radiometry was captured yet. The method is presented in chapter 6.

1.3 Publications

In the course of my thesis work, the following publications emerged:

1.3.1 Peer reviewed abstract and oral presentations

2015:
"A scanner-prototype for geometric and radiometric reconstruction of rock-art sites", Höll Thomas, Holler Gert and Pinz Axel, Computer Application and Quantitative Methods in Archaeology

1.3.2 Peer reviewed papers

2014:
"A novel high accuracy 3D scanning device for rock-art sites", Höll Thomas, Holler Gert and Pinz Axel, ISPRS Technical Commission V Symposium, Oral presentation

2015:
"Cultural Heritage Acquisition: Geometry-Based Radiometry in the Wild", Höll Thomas and Pinz Axel, International Conference on 3D Vision, Poster

2016:
"Radiometry propagation to large 3D point clouds from sparsely sampled ground truth", Höll Thomas and Pinz Axel, Asian Conference on Computer Vision (Workshop), Poster

2

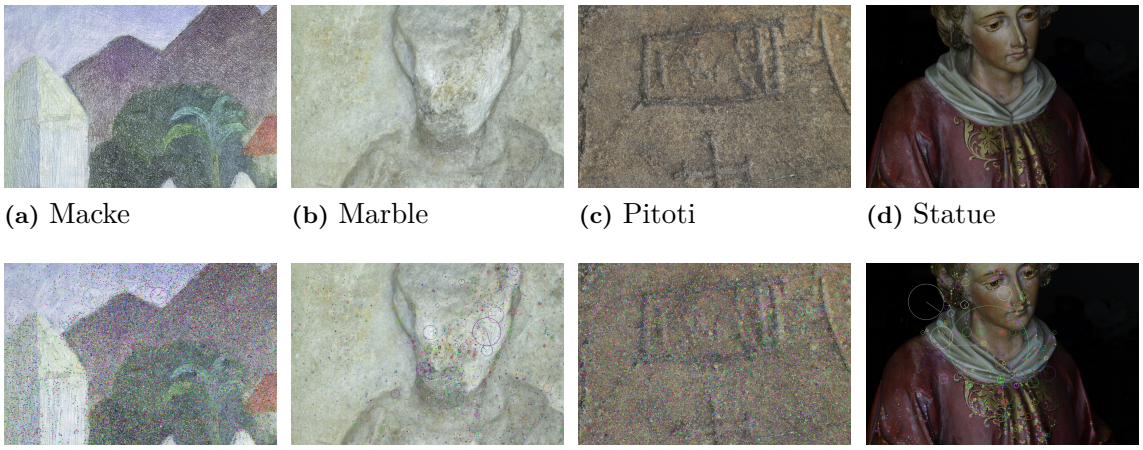
Related work

My task is the 3D reconstruction of surface geometry and radiometric surface properties under the assumption of Lambertian surface reflectance and richly textured surfaces. Under these assumptions, image based reconstruction is the method of choice.

This chapter states how my thesis is related to previous work done in the fields of 3D reconstruction from images and camera based radiometry estimation. W.r.t. Radiometry I also have to discuss related work regarding photo-texture, reflectance functions and shadows. I also briefly look at related work in the context of cultural heritage where various successful methods have been used.

2.1 Geometry

The term “Shape from X” [145] subsumes a variety of methods to infer a scenes geometry like “Shape from Shading” [76], “Shape from Stereo”, or “Photometric Stereo” [165] to name a few. The preferred method I use within the 3D Pitoti project and the thesis is “Structure from Motion” (SfM) [154]. This has several reasons: First, the scanned objects are rich of visual features as can be seen in figure 2.1 where some of the captured images along with extracted visual features are depicted. Second, one criterion for the scanner prototype was affordability and structure from motion is a method that allows to use off-the-shelf cameras and hence avoids the need for custom built solutions. Conducted user studies with potential stakeholders also suggest that the scanner should be lightweight. This is an further argument in favour of SfM.
The next sections state the related work in this field. A good source to get an idea of the historical development in the field of geometric computer vision is given by Sturm [147].



(a) Macke **(b)** Marble **(c)** Pitoti **(d)** Statue

(e) $n_f = 26350$ **(f)** $n_f = 2708$ **(g)** $n_f = 21219$ **(h)** $n_f = 1222$

Figure 2.1: This figure shows examples of four different objects scanned with the scanner prototype and extracted visual features. The extracted features (SIFT [89]) are superimposed on the images of the bottom row. n_f states the number of extracted features. All images are of the same size of 1296×864 pixels. One can see that especially the image “Pitoti” is feature rich.

2.1.1 Epipolar geometry

The epipolar geometry describes the relationship between two different views of a scene. If the two views observe the same 3D point $\boldsymbol{p}$ the two corresponding image points $\boldsymbol{p'^1}$ and $\boldsymbol{p'^2}$ must obey $\boldsymbol{p'^2}^T \boldsymbol{F} \boldsymbol{p'^1} = 0$ which is the result of the co-planarity of the camera centres,

the image points and the 3D point. The matrix $\boldsymbol{F}$ is the well known *fundamental matrix* which is a 3×3 rank 2 matrix. One implication of this constraint is that knowing *e.g.* $\boldsymbol{p'^1}$ and $\boldsymbol{F}$ one can restrict the location of $\boldsymbol{p'^2}$ in the second view to the corresponding *epipolar line* despite the 3D point is unknown.
According to Sturm [147] and Zhang [175] the earliest work on epipolar geometry was done in the 1800's by Hauck [70], Hesse [71], and Sturm [148]. Hesse [71] proposed an algebraic solution that is essentially equivalent to the 7-point algorithm (Hartley [67] and Torr *et al.* [151]) for determining the fundamental matrix, but the link to the epipolar geometry was missing. Sturm [148] and Hesse [71] also stated that there may exist three solutions given seven point correspondences.
The concept of the fundamental matrix was introduced by Faugeras [39] and Hartley [66]. Since then researchers have investigated various methods to estimate the fundamental matrix. The minimal number of point correspondence is 7 and Hartley [67] and Torr *et al.* [151] used the 7-point algorithm to calculate the fundamental matrix.
Using more than 7 point matches ensures a unique solution. An 8-point algorithm was introduced by Longuet-Higgins [87], but its applicability in practice was just since Hartely [68] presented a normalization method to ensure a numerically robust solution.
The literature (*e.g.* [69], [175]) suggests to perform a subsequent non-linear optimization which is based on a parametrization of the rank 2 fundamental matrix and the minimization of a geometric quantity (*e.g.* the distance of points to their corresponding epipolar lines or their estimated re-projections).

2.1.2 Projective reconstruction

Projective reconstruction is the result if nothing about the camera or the scene is known [69]. More precisely, for a real camera, the task is "oriented projective reconstruction" [161], constraining the problem to allow only points in front of the camera. The reconstruction is done up to a global projective transformation. The actual world points and camera locations, orientations and instriniscs are distorted w.r.t the Euclidean space and thus measures of length or angles are pointless. In 1897 Finsterwalder [45] stated that one can obtain a projective 3D reconstruction from a set of uncalibrated images. In his article he also published an algorithm for the case of two images. In the 1990's the topic was rediscovered in computer vision by Faugeras [39] and Hartley *et al.* [65] who state that a projective reconstruction is possible given the fundamental matrix.

2.1.3 Camera calibration and self-calibration

To obtain a similarity or metric reconstruction one needs further information about the camera or scene. This could be achieved by calibrating the camera, the usage of knowledge

about scene (*e.g.* ground reference points), or camera self-calibration.

Camera calibration: For the topic of camera calibration, the survey of Zhang [177] is a good starting point. In practice one can use off-the-shelf software [21] to calibrate a camera based on a planar calibration target [176].

Camera self-calibration: Finsterwalder [45] showed already that camera self-calibration based on the absolute conic is possible but didn't find an analytical solution. Self-calibration based on the absolute conic use the fact that the image of the absolute conic (IAC) depends only on the camera internal calibration this means that the IAC is invariant under Euclidean transformation. The historical first method on self-calibration was presented by Faugeras *et al.* [40] and Maybank and Faugeras [97] and is based on the IAC and the *Kruppa*-equations. It assumes a camera with fixed intrinsic parameters. Later, camera self-calibration was extended towards multiple cameras with varying intrinsic parameters ([72], [122]).

Scene knowledge: Only similarity reconstruction up to unknown scale can be accomplished with a calibrated camera. To calculate the Euclidean structure of the scene and camera positions from a projective reconstruction, Hartley *et al.* [65] use 8 ground reference points. Additional to ground reference points, parallelism of lines and knowledge about vertical planes could be incorporated as constraints to obtain a Euclidean reconstruction [101].

2.1.4 Structure from motion

Structure from motion (SfM) aims at reconstructing the geometry of a scene and the cameras who capture it. This could be a single camera which moves through the scene or a set of individual cameras. In recent years much effort has been spent in developing automatic structure from motion methods which are able to perform the reconstruction task automatically without or with only minor user interaction. Despite the vast amount of work on this topic, some common steps remain the same which are ([38], [121], [52]): feature extraction and matching, recovery of camera pose and scene geometry, and refinement of the geometry.

Feature extraction and feature matching (relating images): Most structure from motion methods depend on interest point detection and matching (an exception is *e.g.* [112], where the autors use 3D curves to refine the camera position). Moravec [102] proposed one

of the first signal based interest point detectors which is based on the *auto-correlation*[1] of an image. He considers the grey level difference of a window and the window shifted in the four directions parallel to the row and column of the image. Based on the minimum of the sum of squared differences he decides if an interest point was found. Several other proposed interest point detectors build also on the auto-correlation approach ([90], [49], [51], [64], [50], [150], [141]) but reformulate it and use the characteristics of the *auto-correlation matrix*

$$\boldsymbol{A} = k * \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} , \tag{2.1}$$

where k is a convolution kernel, I_x and I_y are the image gradients in x and y direction. Shi and Tomasi [141] propose that maxima in the smaller eigenvalue of $\boldsymbol{A}$ are *good features to track*. The famous Harris corner detector [64] detects interest points if $\boldsymbol{A}$ has two significant eigenvalues and demonstrats an excellent performance w.r.t. the criteria repeatability and information content (see Schmid *et al.* [136]).
One major concern regarding interest point detectors is their scale invariance. Lindeberg [85] proposed extrema of the *Laplacian of Gaussian* (LoG) to detect an interest point location and select its representative scale. Based on these findings, Mikolajczyk and Schmid [98] extend the Harris detector to be scale and affine invariant.

In computer vision, the *scale-invariant feature transform* (SIFT) [89] is a widely known and popular combination of interest point detector and descriptor. To detect SIFT keypoints (spatial location and appropriate scale) the algorithm first finds maxima/minima in the Difference of Gaussians (DoG) $D(x, y, \sigma)$ scale space function. This function is constructed very efficiently using a set of octaves. In each octave, a set of Gaussian images is calculated by convolving the input image with Gaussians, each with a specific scale factor. Next, DoG's for a specific scale are calculated by subtracting adjacent Gaussian images. For the subsequent octave the Gaussian images are sub sampled and the process is repeated. The extrema are detected by comparing a pixel to its neighbours, both in scale and space. The location (x, y) and scale (σ) of the detected keypoints are further refined by fitting a quadratic function to the data surrounding the keypoints. In a next step, low contrast keypoints are rejected to increase the stability w.r.t. image noise. The curvature of the keypoints is used to reject detected extrema that correspond to edges. This is done by computing the Hessian matrix $\boldsymbol{H}_D$ of the the corresponding DoG and rejecting all keypoints for which $Tr(\boldsymbol{H}_D)^2/Det(\boldsymbol{H}_D) < (r+1)^2/r$ for $r = 10$ in the paper [89].
Because SIFT features are such an essential part in alot of computer vision processing pipelines, fast computing of these features is essential. Changchang Wu's *SiftGPU* [7]

[1] Auto-correlation uses the *product* of two functions. The correlation used by Moravec based on *squared differences*.

utilize modern graphics card to speed up the SIFT computation.

To handle scenarios in which the distance of two corresponding cameras is large (*e.g.* wide-baseline stereo) one has to find interest points which are at least affine invariant. One such interest point detector was presented by Mikolajczyk and Schmid [98] and is based on the Harris detector. Detected Harris points that do not correspond to local extrema in the LoG scale-space are rejected (scale selection). For each remaining keypoint an affine coordinate frame is calculated based on the second moment matrix.
Another important affine-invariant region detector is the *Maximally stable extremal regions* (MSER) [94] detector. The MSER detector calculates binary regions by thresholding the input images with all possible gray values ($0 \dots 255$). While thresholding the image the algorithm keeps track of the resulting region areas. Regions whose change in area is below a certain threshold are defined as *maximal stable*. Nistér and Stewénius proposed an efficient implementation of the MSER detector [111]. If necessary one can determine an affine coordinate frame for each detected region using *e.g.* its second moment matrix.
A different approach was taken by Mustafa *et al.* [106] in which the authors oversegment an image and use the interesction of three or more regions as interest points. This provides feature points that can be consistently detected across a wide-baseline scenario.

A comparison of various affine region detectors can be found in [100]. In the survey of Tuytelaars and Mikolajczyk [153] the authors present an overview of different invariant interest point detectors.

Once interest points have been extracted, meaningful descriptors are required for subsequent comparison and matching. This *feature description* should be robust to find corresponding interest points in a set of images despite of rotation, scale and illumination changes. If one deals with video sequences [141] or rectified stereo-images [88] one can use metrics like the *sum of squared difference* [63].
In practice, one of the most frequently used image descriptors is SIFT [89]. For a detected keypoint it uses the gradient information in a 16×16 window to calculate a 128 dimensional feature vector. Performance evaluation of different feature descriptors [99] states the excellent performance of the SIFT feature descriptor. The evaluation of Gauglitz *et al.* [54] states results in the context of visual tracking which included usecases which are also resonable in the context of SfM, like *motion blur*. A more up to date discussion on feature descriptors can be found at the *ECCV 2016 workshop on local features* [95] where a new benchmark ([3]) and evaluation metric was presented.

After describing features one needs to find the corresponding features in other images.

This process is called *feature matching*. A feature like SIFT is described by a 128 dimensional vector. Assuming that the Euclidean distance between feature vectors is directly related to their similarity, finding the corresponding feature in another view can be cast as a search problem that finds the nearest neighbour in a high-dimensional space (in the case of SIFT, a 128 dimensional space). In practice one can increase the matching performance if also the distance to the second nearest neighbour is taken into account. This is done by calculating the *nearest neighbour distance ratio* (NNDR) [99] as $NNDR = d_1/d_2$, where d_1 and d_2 are the distances to the nearest and second nearest neighbour. Matches are rejected, if the NNDR is above a certain threshold. The intuition behind this is, that one wants to match feature points that have a distinct nearest neighbour and reject those feature points that match to multiple other points which is the case *e.g.* at repetitive structures. An extensive evaluation of different local feature descriptors and matching strategies can be found in [99].
Feature matching is a very time consuming task and allot of research has been spent into optimizing and speeding up this process. Although some matching strategies are based on nearest neighbour search it is still a computationally intensive task because of the possible large number of features and the high-dimensional search space.
Approximate nearest neighbour (ANN) search methods improve the speed of feature matching. Methods like FLANN [105] are based on *KD-tress* [19] which hierarchically partition the space. FLANN refers to a collection of different ANN algorithms and is able to automatically choose the best algorithm and parameters for the dataset at hand. A GPU accelerated ANN algorithm was proposed by Wieschollek *et al.* [163] which is able to find the (approximated) nearest neighbour of a SIFT descriptor in a dataset with *1 million* entries [78] in *0.02 ms*.
If high-level image information is available, like a semantic labelling of the image, Kobyshev *et al.* [79] present an approach to incorporate this knowledge for feature matching. The benefit is that one can reduce the search space. If, *e.g.* one wants to find features on a "sidewalk", one needs not to search the entire image but just the parts of the image with the semantic labelling "sidewalk". Another benefit of this method is that one can exclude features of the scene which correspond to unreliable content w.r.t. the final SfM reconstruction, like moving objects.
For extremely large image databases, Nistér and Stewénius [110] propose data structure and ideas from document retrieval. Their proposed *vocabulary tree* is used by Hoppe *et al.* [75] for their online SfM to find camera images that possibly contain similar image content. This improves the speed of relating images because one does not need to search every image to find corresponding points but just these images, that are proposed by the vocabulary tree. Cao and Snavely [23] approach this task by learning a classifier on image

pairs to predict which image pair will match or not.

Camera pose and scene geometry: Once point correspondences have been established, the camera pose and scene geometry can be computed using the knowledge of Epipolar geometry, projective reconstruction, and the camera calibration.
To do so, a variety of different methods have been proposed. One commonality of the different approaches is that they depend on robust methods to reject potential outliers *e.g.* wrong corresponding points. One of these methods is the widely used and wellknown *Random Sample Consensus* (RANSAC) [46] method which is used to fit a model (*e.g.* the epipolar geometry) to available data (*e.g.* point correspondences). RANSAC uses as many randomly selected data points as at least are necessary to solve the fitting task, performs the fitting, and checks how many points of the whole dataset are consistent with the resulting model. In this way RANSAC can discard outliers because a model-fit with outliers will not be consistent with the remaining dataset.
Capturing the scene with a calibrated camera results in a *similarity* SfM reconstruction of the scene where the actual scale is still unknown. As already mentioned in section 2.1.3, incorporating scene knowledge can be used to determine the true scale. As scene knowledge, one can use *e.g.* known distances of points in the scene or, if available, the distance between two images.

Refinement: To refine the obtained geometry, bundle adjustment [152] is applied to the estimated scene geometry and camera pose. Bundle adjustment minimizes the reprojection error in every image by modifying the estimated geometry and camera parameters.
Software that can be used for this step includes *e.g.* Bundler [143]´, the Ceres solver [12], and "multi core bundle adjustment" [167].

Structure from motion pipelines and software: The huge amount of work done by the computer vision and photogrammetry community has resulted in the practical applicability of structure from motion methods and pipelines. Papers like [11], [143], [57], or [166] show that one only needs images as input to such pipelines to compute a 3D reconstruction. The pipelines estimate the camera parameters, camera pose and scene structure based on the input images solely.
The usual output of such an SfM pipeline is a collection of camera poses and sparse surface points (corresponding to the detected and matched features). A dense surface reconstruction is often done in a separate stage. SURE [132] and PMVS [53] use the sparse output of a preceding structure from motion reconstruction and the images to reconstruct

a dense surface model. Open source packages like the *Multi-View Environment* (MVE) [4] consist of a software package that allows end-to-end image-based geometric reconstruction that results in a dense surface model. Other freely available SfM reconstruction software includes Bundler [143] (source code), VisualSFM [166] (binary), or openMVG [5] (source code).

2.2 Radiometry

To capture the appearance of an object or of a scene, the geometry alone is not enough. This requires to also capture how the object / scene interacts with light - its radiometric properties.
While capturing the geometry is already a mature topic, capturing radiometry is still a research field. A comprehensive overview of the topic of *appearance acquisition and representation* is given by Weyrich *et al.* [162]. The main topic of this survey is the acquisition and representation of the *Bidirectional Reflectance Distribution Function* (BRDF) [113] and its simplifications and generalizations (*e.g.* anisotropic reflections or subsurface scattering [59]).

2.2.1 Challenges to overcome

Cultural heritage has an inherent interest in preserving the appearance of cultural properties. The *Parthenon* project [33] for example digitized a large-scale scene with the aim to capture also its radiometric properties beyond photo-texture. In case of the reflectance property, the authors sampled the BRDF of a reflectance-characteristic area. The scanning of the BRDF was done during night to eliminate the influence of the daylight. This illustrates one problem or challenge to overcome if the acquisition is done "in the wild": eliminating the incident light. The incident light is necessary to capture an image, but for the sake of appearance acquisition one wants the reflectance property of the object alone.

Eliminating incident light: There are several approaches to tackle this:

- Controlled environment: These approaches perform their measurements under laboratory conditions which means that ambient light can be avoided. Another characteristic of these works is that the reflectance properties are captured with custom built devices.
  Ward [158] uses a helf-sphere half-mirror and a camera with a fish-eye lens to capture surface reflectance. Surface reflectance is captured by the approach of Dana [30] with an off-axis parabolic mirror. Ghosh *et al.* [58] propose to us a special

designed mirror and illumination pattern to acquire the BRDF of a sample. These are three examples that use mirrors to capture the angular dependencies of surface reflectance.

Another approach is to construct specialized hardware with multiple cameras and light sources for capturing the angular dependence, as done for example in the *3D Coform* project [1] with the *DOME* - Hardware [137].

- Shrouding of the incident light: While under laboratory conditions one is able to control or eliminate the influence of incident light, in the case of outdoor work this is not possible. There exist sites where daylight is naturally blocked, *e.g.* caves. For example, the *Rovina*-project [6] has the aim to reconstruct catacombs in which no daylight is present. When this setting can not be achieved, one can block daylight by shrouding with *e.g.* a tent.
- Capture the incident light: Using a *high dynamic range* (HDR) image capture method [34] the incident illumination can be captured and stored in an *environment map* or *light probe*. Such light probes are used by Debevec [31] and Debevec *et al.* [32] to illuminate synthetic and captured objects.
- Estimating the incident light: Some authors propose to estimate the incident light together with the surface reflectance. Romeiro and Zickler [130] estimate jointly the surface reflectance and incident light, given an image and the shape of an object. Similar work to achieve this has been done by Lombardi and Nishino [86].

Another challenge is that the BRDF is a high-dimensional function. The BRDF describes how light is reflected from a single surface point, depending on the incoming light direction and the viewpoint (four dimensions are necessary to describe this). Because an object is often composed of different materials one needs to capture its spatial dependencies. The function that describes this is called *Spatially Varying Bidirectional Reflectance Distribution Function* (SVBRDF) (a six-dimensional function). The following paragraph discusses related work to infer (SV)-BRDF with a sparse set of measurements. The sparseness in the presented work means that only a few images / views or angular dependent measurements are required.

Sampling a high-dimensional function: Measurement of surface properties is an active field in Computer Vision, e.g. Chandraker [25, 26, 24] has investigated a principled tripartite framework underpinning the possibilities and limitations of BRDF measurements with respect to light source, object and camera motion.

An early attempt on surface reflectance estimation was done by Dror *et al.* [131] who treat this problem as a classification task. They motivate their work with the observation that natural illumination has statistical regularities. These regularities interact with the

reflectance property of an object and will cause certain statistical features of the image. The authors use theses features to classify the reflectance property of the depicted object. Zickler *et al.* [180] tackle reflectance estimation as a scattered-data interpolation problem in a mixed spatial and angular domain. They also state that the angular accuracy can be increased for a decrease in the spatial resolution. Their method requires as input a known geometry, known viewpoint, and a point-light-source.
Zhou *et al.* [178] present in their work a method to capture real-world surface reflectance from a small number of views. Given the geometry of an object and an environment map that captures the illumination, their method is able to infer the reflectance properties of the object. Moreover, the authors also analyze how the geometry and lighting effects the total number of views required to estimate the reflectance property of the object. One outcome of their analysis is that fewer views are required for an environmental lighting than as for directional lighting. For the case of directional lighting one can decrease the number of required views by carefully choosing the light/view direction. They conclude that environmental lighting (*e.g.* a natural scene) is superior to directorial lighting for reflectance estimation. This conclusion is supported by the work of Fleming *et al.* [47] who conducted a study on how humans perform in estimating the reflectance properties (*e.g.* gloss) of an object given an image. Another outcome is that for a curved object one needs fewer views to infer its reflectance than for a flat object.
If one has access to a *gonioreflectometer*, one can use the approach of Filip *et al.* [43]. Based on the analysis of a densely sampled BRDF dataset , the authors propose a method that interpolates a BRDF from a sparse set of samples (*e.g.* 2304 samples instead of 518400)
Simplified BRDF representations, such as bivariate BRDF [129], have been used e.g. by [56] to learn a BRDF from sparse measurements.
The research on BRDF profits a lot from available datasets of scanned materials. Matusik *et al.* [96] published a set of 100 scanned isotropic BRDFs and Filip [42] captured 150 anisotropic materials. Such datasets are used to evaluate BRDF models [108], find simplified representations ([129], [109]), or to define priors for reflectance ([130], [86]).

2.2.2 Feasible solutions:

The scientific community in cultural heritage developed feasible solutions to compute accurate geometric reconstructions with highly detailed surface textures.

Color registration and mapping: In cultural heritage acquisition (*e.g. The Digital Michelangelo Project* [8]) it is often the case that one obtains a dense 3D geometry (*e.g.* by laser range scanning) and a set of high quality images. The question is now how to fuse those two kinds of data sources. This is covered under the term *color registration and*

mapping [84]. A two step approach is performed to do so: first, register the image to the 3D geometry and second, map the color from the image onto the dense surface. A comprehensive survey on this topic can be found in [118].
The registration can be performed using various methods. A straight forward method is to let a user select corresponding points on the geometry and the images. While this is a reasonable approach [35] it does not scale well with the number of images. If a camera is attached to the laser scanner, one can use this information to register the acquired images to the 3D geometry automatically *e.g.* [124]. Feature based methods extract 3D (from the geometry) and 2D (from the images) features and try to match them. These methods are good for man-made environments. If radiometric information from the laser scan is available (*e.g.* intensity of the reflected beam), Sequeira and Goncalves [139] use the edge information of the laser and image data to fuse them. Another idea is to maximize the mutual information [156] between the dense 3D data and the image intensities to register both. The intuition behind this is that the observed image intensities and the surface normals (obtained by the dense 3D data) are somehow correlated and by maximizing the mutual information one wants to find a set of parameters that explain both data entities. The most promising methods to automate the registration process are based on a *structure from motion* (SfM) approach. The initial problem of registering 2D to 3D data is split up into two steps: first, the 2D correspondence problem between the images is solved and a sparse 3D geometry is reconstructed. This sparse 3D geometry is subsequently registered with the dense 3D geometry obtained by the laser scanner which can be done using robust methods *e.g.* RANSAC [46].
After the registration of 2D image and 3D geometry, the color information of the images must be mapped onto the surface. This is done by projecting and blending the color information of the images. The weights for blending could be computed based on the geometry, image sharpness, orthogonality of the image w.r.t. the surface normal or distance of the object to the image origin. Blending is a simple and fast operation but is prone to blurring the color information. To overcome this, Dellepiane *et al.* [36] select a leading image and wrap the other images via optical flow to this image before the content is mapped onto the surface. As a result the obtained mapping exhibits more sharp details.

Polynomial texture map: Another direction in appearance acquisition in the context of cultural heritage is the *polynomial texture map* (PTM) [93]. For PTM, one acquires a set of images under varying illumination direction and models the object in image space were one pixel encodes the usual color information (R, G, B) and additionally the illumination direction.

2.2.3 Color transfer

Color transfer allows the user in a simple way to change the color appearance or characteristic of an image. Given a source and a target (reference) image, the aim is that the color charactersistic of the source image matches that one of the target (reference) image. A comprehensive survey on the topic of color transfer can be found in Faridul *et al.* [37]. The authors categorize color transfer algorithms into three main classes: geometry-based, statistical and user-assisted methods.

- Geometry-based methods: For some applications like image stitching or multi-view image capture it is preferable that the captured set of images have the same color characteristic (appearance). For example, a depicted object should have the same appearance in all captured images. This is not necessarily the case if *e.g.* two different cameras are used.

  Geometry-based methods find correspondences between the images and ensure that *e.g.* a particular object has the same color in all images where it is depicted. This can be achieved by using sparse correspondences between images. Yamamoto *et al.* [172] are using this approach to correct the color of multiple images. First, they select a reference camera (all other images are mapped to the color characteristic of this camera). In a second step, they are detecting correspondences between images using SIFT [89] features. Third, from these correspondences they calculate a lookup table that maps colors from the target image to the reference image. Oskam *et al.* [114] also use also correspondences between images. But instead of treating the color transformation as a lookup table, they describe the transformation as a vector field. The authors handle colors as 3D points in *Lab*-color space. For each corresponding pair they calculate a translation vector that maps the image color to the reference color. For parts in the Lab-color space where no correspondence information exists, the authors interpolate the transformation vector using radial basis functions. The result is a vector field that transforms the entire color gamut of the images to the reference image.

  Instead of sparse correspondences one could use corresponding regions or dense correspondences. The latter case is used by HaCohen *et al.* [62]. The authors use the *Generalized PatchMatch* algorithm [17] to find densely corresponding regions. The proposed method is an iterative process in which the following steps are performed for different scales (coarse-to-fine approach): First, for each patch in the source image find a nearest neighbor in the reference image. Second, aggregate consistent regions of matches. Third, fit and apply global color transfer, and as a last step adjust the search range for the next iteration.

  In addition to the presented methods so far, one could also use the semantics of the

image content. For example, if faces appear in the source and target image it would be beneficial if those parts of the images are used to estimate a color transfer. Work in this direction was done by Wu *et al.* [168] and Yang *et al.* [173].
Another idea is to use salient maps to guide the color transfer between images. Xia [170] divides the image into two regions based on salient maps. Next, the author uses the information by the corresponding salient regions to estimate the color transfer.

- Statistical methods: The color value of an image pixel is a point within a 3-dimensional color space. The color appearance or style of an image is characterized by the distribution of these points in 3D. In that sense, the essence of color transfer is to reshape the color distribution of the source image so that it is equal (as good as possible) to the distribution of the reference image. Statistical methods use this idea of reshaping distributions and do not rely on correspondence information between images.
Usually the color channels of an image in *RGB*-color space are highly correlated. This means that changing the color distribution of one channel will affect the other channels as well and hence channel independent color manipulation will lead to undesirable artifacts.
In the seminal work on color transfer, Reinhard *et al.* [127] use the $l\alpha\beta$-color space [134] that minimizes the correlation between color channels to overcome this. For each color channel I^c in this color space, the color transfer is calculated in a simple and fast way using only the channels mean μ^c and standard deviation σ^c: $I_o^c = \sigma_r^c/\sigma_i^c(I_i^c - \mu_i^c) + \mu_r^c$, where the subscripts o, r, and i correspond to the output, reference and input image. Pouli and Reinhard [123] perform a channel wise *histogram specification* [60] (histogram matching) to match the histograms of the source image and the reference image. They present furthermore a scale-space approach for doing so that enables the user to control the amount of transferring the color appearance. To improve the fidelity of the color transfer, Xiao and Ma [171] propose a method that combines histogram specification and gradient-preserving optimization.
The presented approaches so far rely on a color space that decorrelates the three color channels. The performance of this decorrelation is scene dependend. As Faridul *et al.* [37] pointed out, the $l\alpha\beta$-color space [134] used by Reinhard *et al.* [127] is tailored towards natural images and hence it cannot be guaranteed that it will work on other scene types. Abadpour and Kasaei [10] use *principal component analysis* to decorrelate the color space.
Approaches that can deal with correlated color spaces tackle color transfer as a 3-dimensional problem and perform the transfer not channel wise, instead they transform the 3-dimensional color distribution directly. Pitié *et al.*([119] and [120]) iter-

atively match 1D projections of the 3-dimensional color distribution. For that, the authors randomly select a 3D rotation matrix and project the color distribution of the images onto the axes of this matrix. For each of the resulting 1D projections a histogram specification is performed to match the reference image color distribution. This process is repeated until convergence.
In [126], Rabin *et al.* established a link between this prior work on distribution transfer and the *Wasserstein metric*. The authors also propose the *Sliced Wasserstein Distance* to speedup the computation of the *Optimal Transport* [155] problem which is also applied to solve the color transfer task (Ferradans *et al.* [41], Rabin *et al.* [125]). More details on this are presented in section 6.1 of this thesis.

- User-assisted methods: User-assisted methods are used for various reasons. User defined masks [28] and strokes [160] are used to select regions where no color manipulation should occur. Other methods allow the user to select the corresponding regions manually ([114], [159]). In addition, Levin *et al.* [83] use the color of the user's input stroke to transfer the color.

2.3 Contributions of this thesis w.r.t. geometry and radiometry

1. Geometric reconstruction: As stated in this chapter, SfM with a single calibrated camera will lead to a similarity reconstruction up to scale. To infer the scale of the scene, I use a stereo camera setup. This setup is calibrated and the resulting additional knowledge of the relative orientation (baseline and rotation) between the stereo cameras is used first, to estimate the scale of the scene and second, to guide the optimization during bundle adjustment.
   In addition, I fuse tachymeter measurements and scene structure obtained by SfM to obtain a geo-referenced reconstruction. This fusion step as well as the incorporation of the knowledge about the stereo setup is implemented in a constrained bundle adjustment process.
2. Radiometric reconstruction: The presented approach reconstructs Lambertian radiometric surface properties "in the wild". The term "in the wild" refers to a non-laboratory and non-controllable setting. For this purpose, we developed a portable scanner with a custom built and geometrically calibrated LED light source that captures the reflectance of surfaces even under bright daylight.
   Dense surface reconstruction, estimated camera poses and geometric knowledge of the light source are used to estimate the incident light onto the surface. In a further step, this estimation is used to infer the Lambertian surface reflectance.

I want to stress that the recovered surface property is the Lambertian albedo of the material represented in a certain color space.

3. Radiometry propagation: Besides the estimation of Lambertian surface properties I propose an approach which is able to propagates known radiometric values into areas of a large scene where only photo-textured 3D reconstruction are available.

These three contributions are discussed in detail in chapter 4 (geometry), 5 (radiometry), and 6 (radiometry propagation).

3

The 3D Pitoti Project: method, scanner and data

To better understand the particular contribution of my thesis, this chapter states the broader context in which it arose: the 3D Pitoti project. See [13] for a more detailed description, especially of the acquisition process.

3.1 The 3D Pitoti method

As pointed out already in section 1.1, the aim of the 3D Pitoti project is was to acquire, process and present prehistoric rock-art in its landscape context. This overall aim can be broken down into five sub-goals (directly taken from [13]):

1. General requirements analysis and specification of the 3D Pitoti hard- and software components.
2. 3D recording and registration of petroglyphs, rock-panels and large-scale landscape for selected sites in Valcamonica.
3. Automated segmentation and classification of individual petroglyphs, based on their 3D structure.
4. 3D immersive, multiuser, interactive visualization at various scales.
5. Dissemination to a wide range of key stakeholders, including archaeologists, surveyors, museums, and schools.

The sub-goals (3), (4), and (5) depend on a high-quality geometric and radiometric reconstruction of the rock-art and its landscape.
Because an individual petroglyph is very small compared to the landscape in which it is situated, the project consortium agreed to examine a site at three characteristic scales to be able to describe rock-art in its landscape context:

1. Macro scale (meter): parts of the valley were captured with a motorized glider equipped with a high resolution camera.
2. Mid-range (10^{-2} meter): Rock-panels were scanned with a micro aerial vehicle.
3. Micro (10^{-4} meter): The individual rock-art was scanned by a custom built scanner prototype (see section 3.2).

To achieve high-quality reconstructions over these vastly different scales, the project consortium decided to develop a multi-scale photogrammetry-based solution in which each characteristic scale is reconstructed based on a photogrammetric method and the different scales are registered based on a common reference frame.

Benefits of the 3D Pitoti project are *e.g.* large-scale interactive visualization and analysis as well as seamless transition between reconstructions of vastly different scales. To achieve this one has to overcome challenging tasks like the spatial registration of different reconstructions as well as the "registration" of different radiometric reconstructions. Moreover, non-experts in photogrammetry should be able to do the scanning. A further challenge is that the scanning is done outdoors in a non-controllable environment in which one is confronted with *e.g.* direct sunlight and cast shadows.

3.2 The scanner prototype

An integral part in answering my research questions (section 1.2) and accomplishing the challenges of the 3D Pitoti project is the scanner prototype shown in figure 3.1. It is a

custom built scanner that was designed, built and tested within the 3D Pitoti project and was joint work with my colleagues *Gert Holler*, and *Axel Pinz*. A conducted study [13] with potential end-users of the scanner revealed the following requirements for the scanner (sorted from most important to least important):

- Weight: The scanner should be light weight. A huge number of rock-art sites are in remote places where the scanner must be carried by foot this restrict the weight of the scanner.
- Spatial resolution: As one can see, *e.g.* in figure 1.1, our rock-art of interest exhibits very fine detailed structure that should be preserved.
- Price
- Battery life: The operation time of the scanner should be a working day. We accomplish that with exchangeable batteries.
- Waterproofness: The current prototype is not waterproof, but can be protected with a rain cover.

In addition we require that the scanner should capture rock-art in a fast way and be easy to operate, even from non-experts in the field of photogrammetry or computer vision. A further restriction for our scanning method is that we were not allowed to stick reference markers on the rock surface. This leads to a marker free reconstruction approach.

Figure 3.1 shows the resulting scanner prototype. All components are mounted around a central carbon fibre pole, resting on a carbon fibre mini-tripod. The usage of carbon fibre was mandatory to decrease the weight of the scanner and at the same time ensure its stiffness.

To capture and reconstruct the geometry of the rock-art we chose a *Structure from Motion* (SfM) approach. The reason for this is, first, the rock surface as well as the particular rock-art is rich of visual features and second, SfM fits in the common SfM processing pipeline of the 3D Pitoti project (see page 12 for related work on *SfM*).

We capture the necessary images for the SfM reconstruction with a stereo-camera setup and use the additional information about the stereo baseline to reconstruct the scale of the rock-art. This approach to Euclidean reconstruction was chosen because we were not allowed to stick reference markers to the rock surface and so we had to use the information about the baseline to infer the scale.

We also installed a 360° prism on the scanner to enable the possibility of a direct geo-referenced reconstruction.

A custom made LED-flash consisting of 220 high-power white LEDs is installed to capture the radiometric surface properties. The board is designed in such a way that the intensity of the LED-flash can be controlled. In addition groups of 10 LEDs can be controlled individually to enable a directed illumination of the scene. How this is done in detail will

be explained in chapter 5.
A Raspberry Pi computer controls the acquisition process during scanning. To obtain on-site feedback we transmit the captured images to a tablet PC where a sparse SfM reconstruction [75] is performed online, so that the user knows which part of the scene already has been scanned by a sufficient number of images.

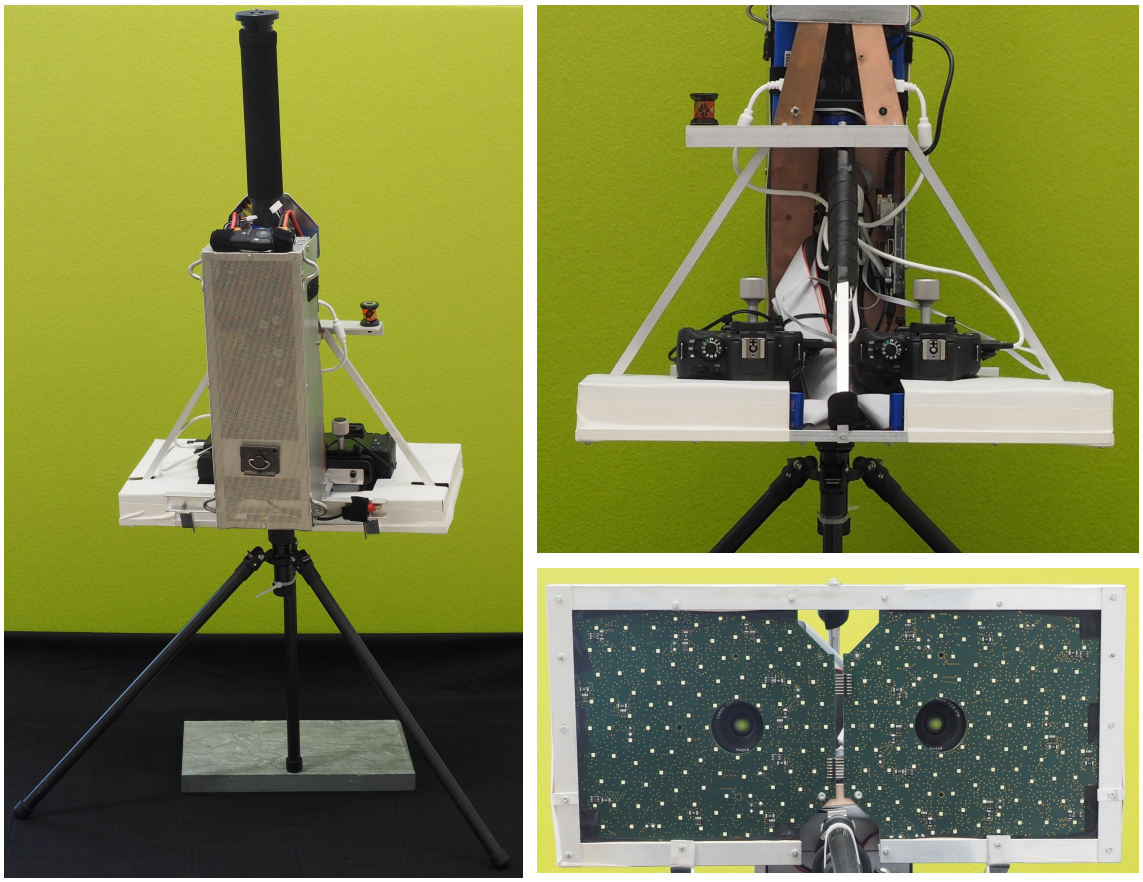


Figure 3.1: Left: Scanner prototype mounted to a carbon pole, resting on a miniature carbon tripod. Top right: detail close-up showing the calibrated rigid configuration of a 360° prism, a stereo rig, and a custom illumination unit. Bottom right: The illumination consisting of 220 high power LEDs. The two cameras look through the circular holes.

The most important advantages of the scanner w.r.t. the project and this thesis are:

- It is able to produce a Euclidean (metric) geometric reconstruction of the scene without additional scene knowledge. The cameras are calibrated and the scale of the scene is obtained via the calibrated stereo-rig.
- Additionally, in co-operation with a tachymeter one can obtain a geo-referenced reconstruction without the need of *e.g.* georeferenced fiducial markers on the rock.
- With the custom illumination it is possible to get a radiometric reconstruction of

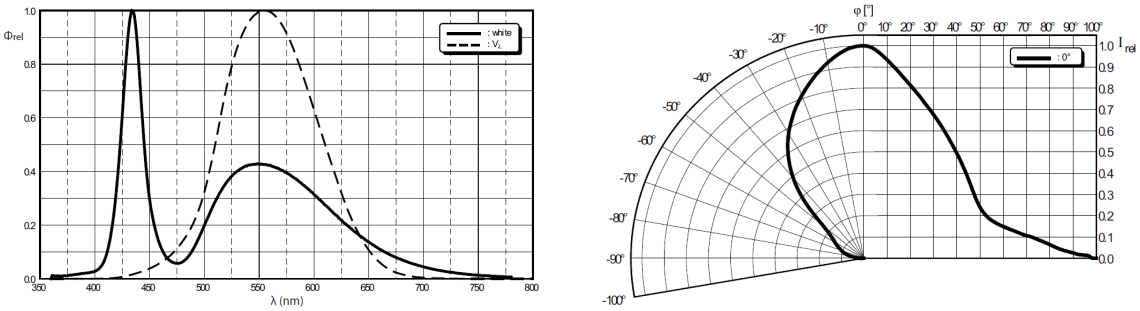
the rock-art without shrouding.

The following paragraphs summarizes the *implementation details and scanner configuration.*

Imaging configuration: We decided to use two *Canon EOS 100D* DSLR [1] cameras for the stereo-camera setup. Each camera is equipped with a *40mm f/2.8* lens. The baseline of the stereo-rig is $162mm$. At a working distance of $47.5cm$ the scanner captures a footprint of $36 \times 14cm$.

In practice, to ensure a large *depth of field*, the aperture of the camera was set to be *f/16* which results in a depth of field of about ±5cm. The exposure time is *1/160* seconds and the *ISO* is 100 for all our captured images. Such low values could be achieved because of the intense light of our LED flash.

Illumination: The custom built LED illumination consists of 220 *Osram CR7P* high-power LEDs. Each LED has a luminous power of $144lm$ (typical). The *relative spectral emission* and *radiation characteristic* are shown in figure 3.2 (directly taken from the datasheet [115]). The size of the illumination box is $47.5 \times 22.5cm$. The total illuminance on ground is $160000lx$.



(a) Relative spectral emission **(b)** Radiation characteristic

Figure 3.2: Relative spectral emission (a) and radiation characteristic (b) of the used *Osram CR7P* LEDs [115]. $V(\lambda)$ in (a) corresponds to the *standard eye response curve.*

Tachymeter: To measure the position of the scanner with a tachymeter we attached a *Leica mini prism* (GRZ101) on the scanner. The distance between the left camera of the stereo-rig to the prism is $35cm$.

[1] Digital single-lens reflex

3.3 Available data within 3D Pitoti

During our field-work we collected a huge amount of data and reconstructions. This section introduces the main dataset that I use within this thesis. Additional data sets which are not related to the 3D Pitot project and show the general applicability of my approach are presented later in the respective experimental sections.
Not all presented data in this section were scanned and reconstructed by me. Ground truth acquisition and geo-referencing of various sites in Valcamonica was performed by project partner *ArcTron* [2]. Micro aerial vehicle based reconstructions as presented in section 3.3.1 were performed by project partner ICG (Institute of Computer Graphics and Vision).

3.3.1 Micro aerial vehicle (MAV) based reconstruction of Seradina I Rock 12C - Ser12c

Figure 3.3 shows the reconstruction of a rock panel called *Seradina I rock 12C*. Within the project consortium we decided that we wanted to use this rock panel as the showcase of the concepts and technology developed within the project.
Project partners (ICG) acquired a total number of 1099 images with an micro aerial vehicle and performed SfM reconstruction resulting in a textured 3D mesh with $372k$ vertices. The rock area (figure 3.3 (A)) covers approximately $400m^2$ with a diameter of about 19m. The geo-referencing was obtained with the help of ground reference targets (figure 3.3 (C)).

3.3.2 3D Prints and Ground Truth

In the early stage of the project, our project partner ArcTron3D took high-quality scans of selected rock panels and rock art using various technical equipment (*e.g.* laser scanner, structured light scanning, high-resolution macro-photography, shrouding of daylight, etc.). We use this ground truth to validate the accuracy of our reconstruction approach.
To allow us experimenting with and testing our scanner in the lab, ArcTron provided us with high-quality 3D prints of the *Camunian rose* and a warrior. Both prints are shown in figure 3.4. The overview of available ground truth data for the rock panel Seradina I rock 12C is shown in figure 3.5a (image taken from [135]). Figure 3.5b shows the ground truth reconstruction of *Area 3*.

[2] `http://www.arctron.de`

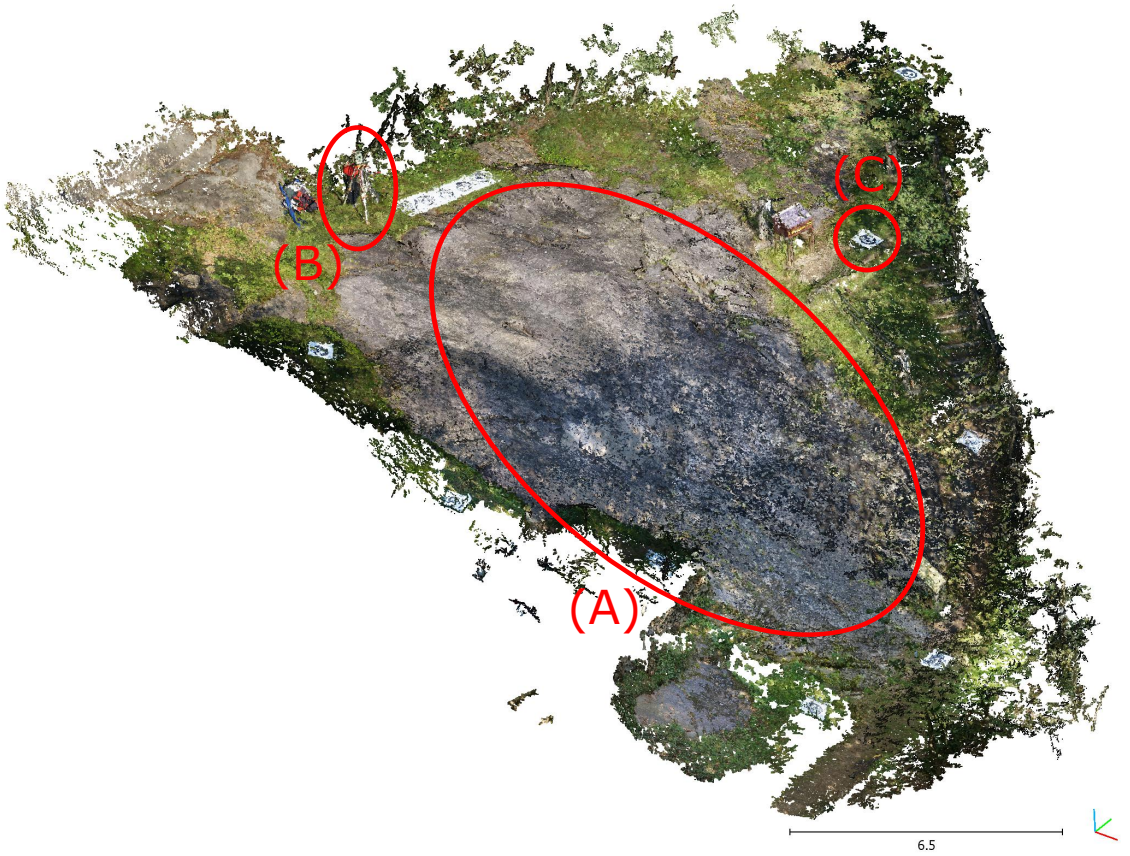
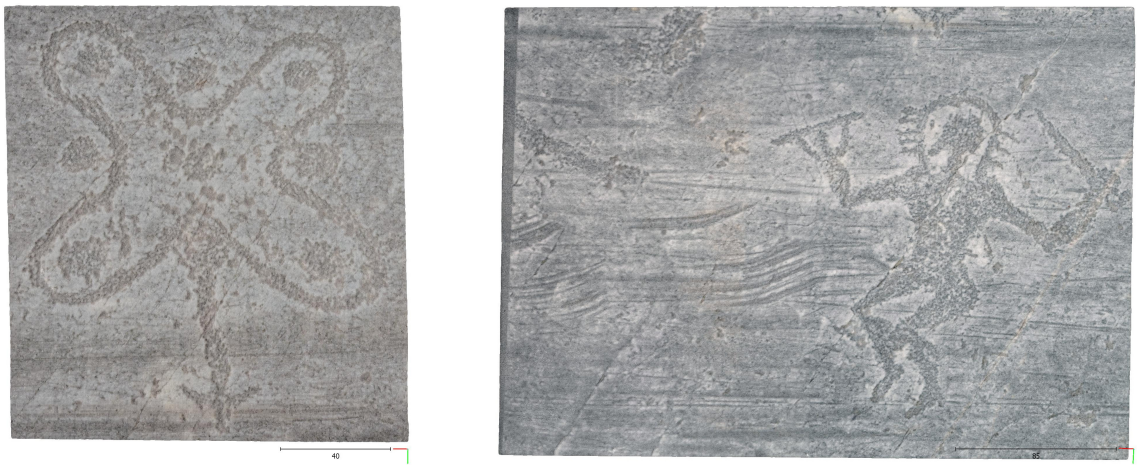


Figure 3.3: This figure shows the point cloud reconstruction of *Seradina I rock 12C* obtained by micro aerial vehicle based image acquisition. The rock area (A) covers approximately $400m^2$. For scale comparison, I also marketed the tachymeter in (B). In (C) one can see a ground reference target (GRT) which is used to geo-reference the final reconstruction. Note, the GRT is not required to geo-reference a reconstruction obtained by the scanner prototype.

3.3.3 Scanner based reconstruction of Areas 3, 4, 5, and 10 - AREA345 and AREA10

Based on the ground truth acquisition done by ArcTron, we scanned several areas of the Seradina I Rock 12C panel. We combined the areas 3, 4, and 5 into the dataset *AREA345* (see figure 3.6) and captured also the elongated scene area 10 (see figure 3.7). The image caption of both figures states the details of the captured datasets.
In both figures (figures 3.6 there and 3.7) are ground reference targets (GRT) visible. To validate the performance of our geo-referencing approach we placed these markers on the rock-panel. This was solely done for validation purpose because our solution works without any artificial markers on the rock.



(a) Camunian rose **(b)** Warrior

Figure 3.4: Two high quality 3D prints provided by our 3D Pitoti project partner ArcTron3D. We use the two prints to validate the accuracy of our 3D reconstruction approach.

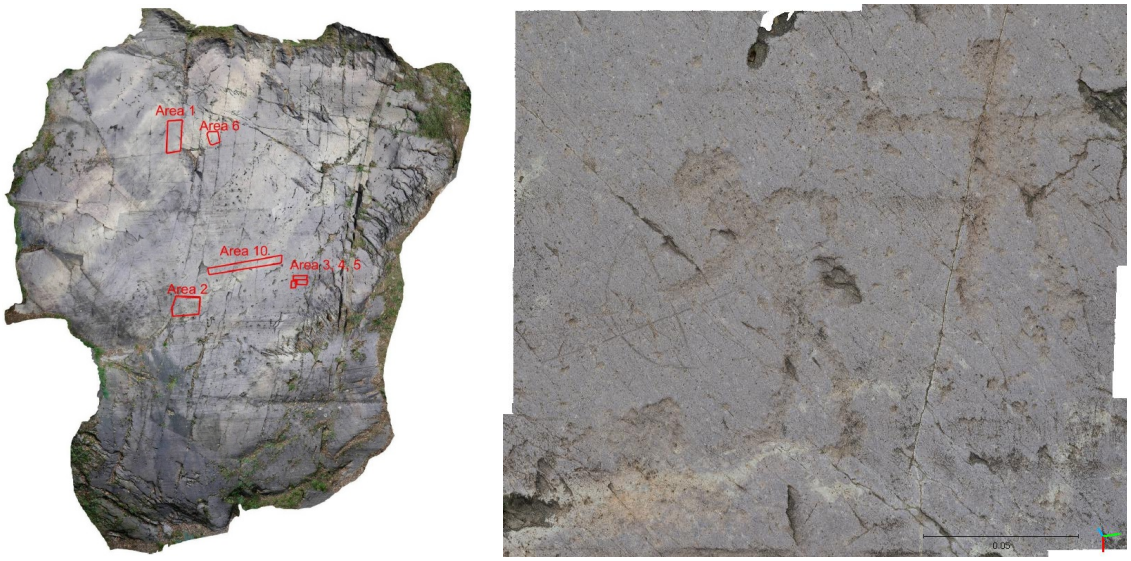


Figure 3.5: (a) shows an overview of available ground truth on the rock panel Seradina I Rock 12C (denoted by "area" followed by a number). (b) shows the ground truth data for area 3.

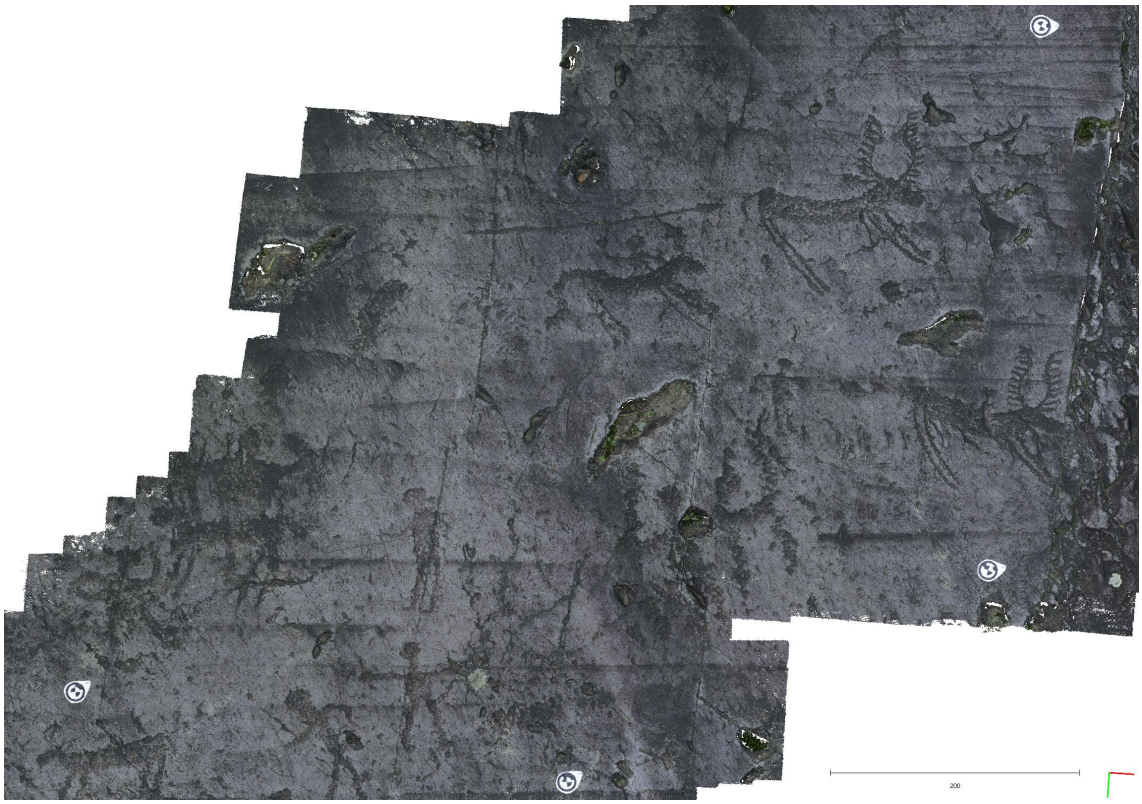


Figure 3.6: This figure shows the scanned areas 3,4,5 of the Seradina I Rock 12C panel. The diameter of the scene is approximately 1 meter and took about 45 minutes of scanning with our scanner prototype. The final reconstruction was obtained using 174 camera images.

Figure 3.7: This figure shows a scan of area 10. The entire area is an elongated structure of approximately 2.6 meter length and was reconstructed using 104 images.

4

Geometric reconstruction

This section describes a method, based on the scanning device (section 3.2), how to reconstruct the 3D geometry of richly textured surfaces. The goal is Euclidean reconstruction at a high spatial resolution better than 0.1mm with or without geo-referencing of the obtained result. Utilizing all available measurements of the scanner we can produce a geo-referenced 3D reconstruction of the sites. This is achieved by augmenting an initial Structure from Motion reconstruction with additional knowledge of the scanner's stereo-setup and tachymeter measurements. The stereo-setup is used to determine the scale of the reconstruction and the tachymeter measurements are used at first hand to geo-reference the reconstruction, and second to compensate the drift induced by the Structure from Motion method. By doing so, a geo-referenced 3D reconstruction is obtained with an accuracy of 0.1mm. A major advantage of this approach is, that the geo-referenced

reconstruction is obtained without the usage of ground control points and that one could easily combine reconstructions of vastly different scales (see for example figure 4.1 [1]). On the other hand, if a geo-referenced reconstruction is not necessary the scanner can also be used without a tachymeter.

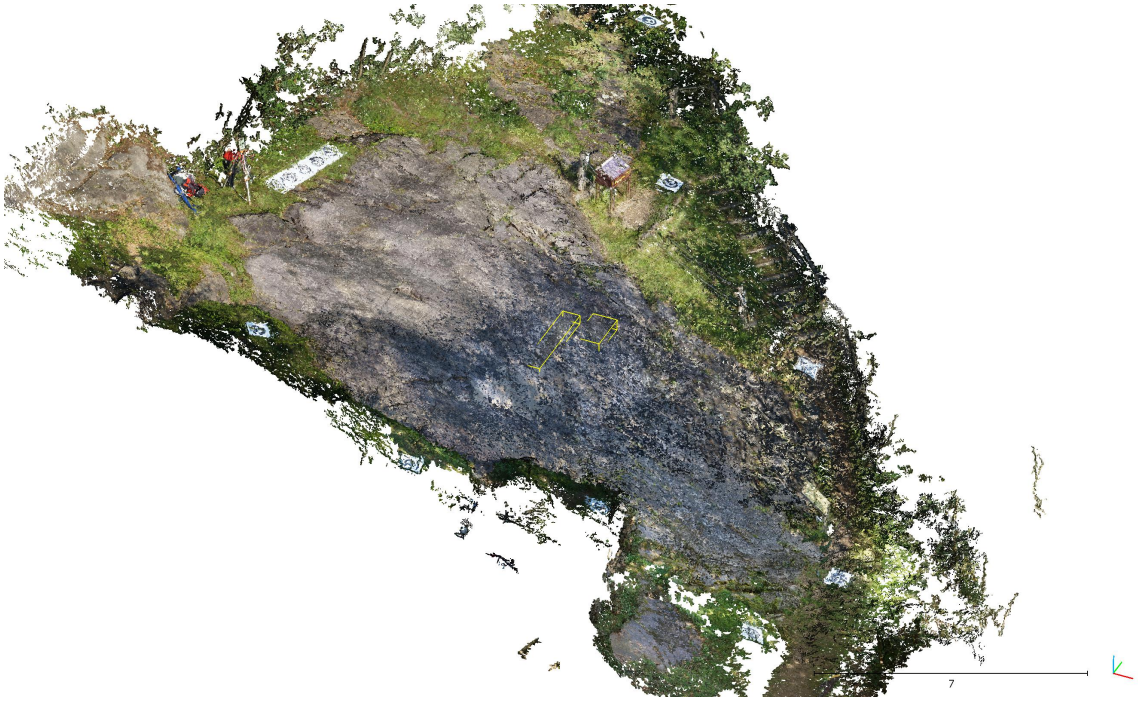


Figure 4.1: Fusion of vastly differently scaled reconstructions. The large reconstruction (presented in figure 3.3) shows the rock panel of Seradina 12C and was obtained using a micro aerial vehicle (MAV) equipped with a camera. The two yellow bounding boxes at the center of the rock-panel indicate two reconstructions obtained with our prototype scanner. The size of the MAV based reconstruction is 1.42GB (372k vertices), the two scanner based reconstructions are 1.5GB (58.9Mio points) and 2.8GB (111.8Mio points). The vastly different scales of the two types of the reconstruction are obvious. The experimental validation of this setting can be found in section 4.3.2.

4.1 Geometric calibration of the scanner

The scanner in fig. 3.1 consists of several components that are rigidly mounted and need geometric calibration. Two cameras and a 360°-prism form a rigid triangular structure, as shown in fig. 4.2. In addition, two custom LED illumination boards with a total of 220 individual high power LED light sources are rigidly attached to the cameras. This

[1] I greatly acknowledge the work of my colleagues at ICG and their kindness to provide me the reconstruction of the rock-panel.

section describes the geometric scanner calibration and related experimental results. First, it introduces the calibration of the stereo rig, second, the estimation of the 360°-prism location w.r.t. the rig, and third, the estimation of the pose of the 220 LEDs w.r.t. the rig.

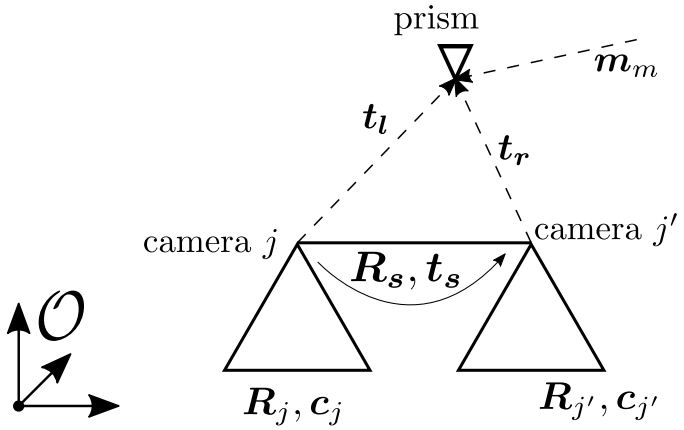


Figure 4.2: The two cameras j and j' of the stereo rig together with the 360° prism form a rigid, triangular structure. The georeferenced prism position $\boldsymbol{m}_m$ can be measured from outside using a tachymeter. All the remaining entities in this figure are estimated by geometric calibration. The relative rotation $\boldsymbol{R_s} \in \mathfrak{so}(3)$ and translation $\boldsymbol{t_s} \in \mathbb{R}^3$ between the stereo camera pair is estimated during the stereo-rig calibration. The relative translation of the prism w.r.t. the camera ($\boldsymbol{t_l} \in \mathbb{R}^3$ and $\boldsymbol{t_r} \in \mathbb{R}^3$) is the result of the prism calibration. The pose of each camera ($\boldsymbol{R_j}$ and $\boldsymbol{t_j}$) is relative to a reference frame origin $\mathcal{O}$ and is estimated by the SfM computation.

4.1.1 Stereo-rig calibration

I first calibrate the stereo-rig using a coordinate measurement machine to acquire a set of $2D \leftrightarrow 3D$ point correspondences with which I estimate the camera projection matrix $\boldsymbol{P}$ for both cameras (using the *Gold Standard algorithm* from Hartley and Zisserman [69], page 181). I use the camera projection matrices as initial values for the subsequent estimation of the lens distortion coefficients and the refinement of the relative orientation between the camera pair, i.e. the rotation matrix $\boldsymbol{R_s} \in \mathfrak{so}(3)$ and translation vector $\boldsymbol{t_s} \in \mathbb{R}^3$. Table 4.1 presents the result of a typical camera calibration.

To validate the stability of the calibration during field work I recorded the calibration parameter before and after the field work. The first calibration was done at the 6th of July 2016. After this calibration the scanner was sent to Italy for field work. After we got the scanner back, a second calibration was done at the 19th of September 2016. The difference in the baseline was 0.105mm, which suggests a largely robust and stable me-

Parameter	Description	Value
Left camera:		
(f_x, f_y)	Focal length in x and y direction in pixel	$(9721.930, 9721.271)$
(p_x, p_y)	Coordinates of the *principal point* in pixel	$(2592, 1728)$
$(k1, k2)$	Radial distortion coefficients	$(-0.075, 0.126)$
$(k3, k4)$	Tangential distortion coefficients	$(2.946 \times 10^{-4}, -6.920 \times 10^{-5})$
Right camera:		
(f_x, f_y)	Focal length in x and y direction in pixel	$(9738.445, 9737.924)$
(p_x, p_y)	Coordinates of the *principal point* in pixel	$(2592, 1728)$
$(k1, k2)$	Radial distortion coefficients	$(-0.078, 0.149)$
$(k3, k4)$	Tangential distortion coefficients	$(-1.121 \times 10^{-4}, -1.993 \times 10^{-4})$
Stereo configuration:		
$\boldsymbol{R_s}$	Rotation of the right camera w.r.t. the left camera	$\begin{bmatrix} 0.999 & -0.002 & 0.025 \\ 0.002 & 0.999 & 0.006 \\ -0.026 & -0.005 & 0.999 \end{bmatrix}$
$\boldsymbol{t_s}$	Translation of the right camera w.r.t. the left camera in mm	$[-162.023, 0.791, 1.010]$

Table 4.1: This table presents the result of a typical stereo-rig calibration. The parameters are according to the parameters used in the *Matlab Camera Calibration Toolbox* [21] or *OpenCV* [22]. For both cameras I fixed the principal point to be the center of the image.

chanical assembly of the scanner.

To further estimate the impact of the varying calibration, I use the two mentioned calibration results and reconstruct the *Camunian rose* 3D print (see figure 3.4). The mean absolute distance between the dense 3D point cloud is 0.045mm.

4.1.2 Estimation of the 360°-prism location w.r.t. stereo-rig

Based on the known stereo-rig calibration, the offset between the 360° prism and the stereo-rig can be estimated. For this purpose, a calibration target was designed which is used to establish corresponding measurements between the scanner and a tachymeter. First, the pose of the calibration target relative to the stereo-rig is estimated and then points on

the calibration target and the position of the 360°-prism are measured with a tachymeter (figure 4.3 shows a sketch of the calibration target and the scanners triangular structure). Given the known geometry of the target, the known pose of the target relative to the stereo-rig and to the tachymeter, and the position of the prism seen by the tachymeter, the position of the 360° prism relative to the two cameras is calculated. $\boldsymbol{t_l} \in \mathbb{R}^3$ denotes the position of the prism relative to the left camera, and $\boldsymbol{t_r} \in \mathbb{R}^3$ denotes the position of the prism relative to the right camera.

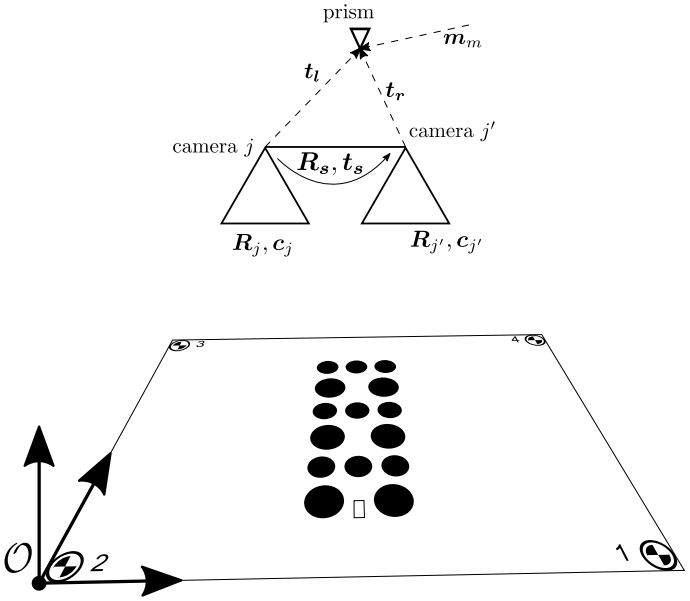


Figure 4.3: Sketch of the proposed calibration method to estimate the 360° prism w.r.t. the stereo-rig. Because the stereo-rig was calibrated beforehand, $\boldsymbol{R_s}$ and $\boldsymbol{t_s}$ are known. First, the pose $\boldsymbol{R_j}, \boldsymbol{t_j}$ and $\boldsymbol{R'_j}, \boldsymbol{t'_j}$ of the stereo-rig relative to the target reference frame $\mathcal{O}$ is estimated. As a result the corner points 1, 2, 3, and 4 relative to the cameras can be calculated. These four points are measured again with a tachymeter including also a measurement of the 360° prism. The knowledge of the corresponding four corner points and the prism measurement is used to estimate $\boldsymbol{t_l}$ and $\boldsymbol{t_r}$, i.e., the position of the prism w.r.t. the stereo-rig.

To validate the accuracy of this approach, the following experiment was conducted: The corners of two Pitoti 3D prints were equipped with fiducial markers, the prints were scanned by the scanner-prototype and reconstructed using the approach presented in this chapter. Next, I calculated the distance between the measured fiducial markers and the reconstructed fiducial markers. The results are presented in table 4.2. The mean distance is $5.52mm$ which lies in the range of the tachymeters accuracy. Figure 4.4 shows the 3D reconstruction of the two 3D prints and the fiducial markers on them.

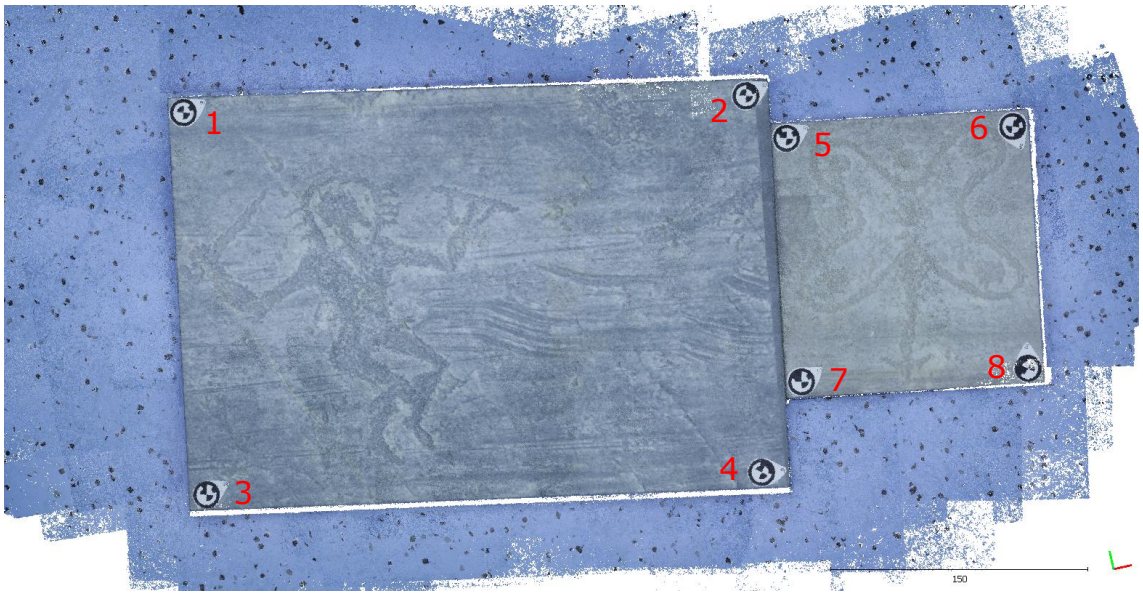


Figure 4.4: This figure shows the reconstruction of two Pitoti 3D prints and fiducial markers at their corners. The red numbers correspond to the ID of the fiducial markers. I estimated the distance between the reconstructed fiducial markers and the tachymeter measurements of them. The mean distance is $5.52mm$. The distance for every fiducial marker is listed in table 4.2.

Fiducial marker:	1	2	3	4	5	6	7	8	*mean*:
Distance in *mm*:	4.05	4.35	5.83	6.65	4.63	5.97	5.61	7.10	5.52

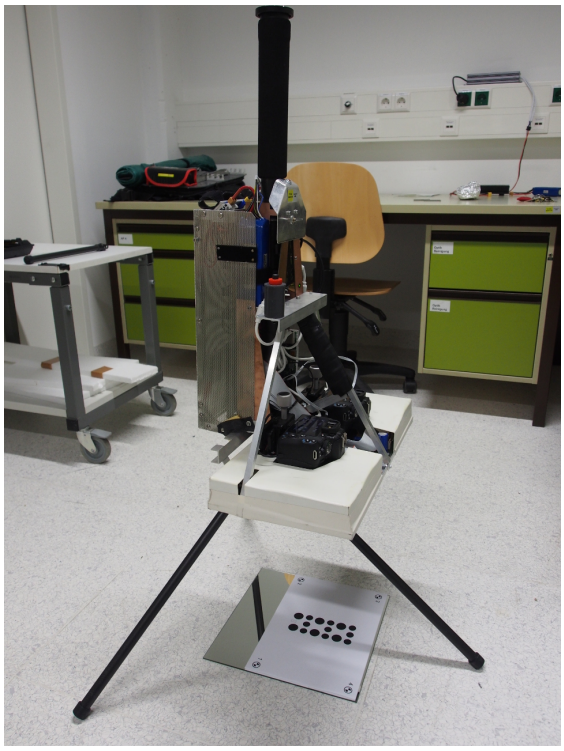
Table 4.2: Distances between tachymeter measurement and scanner reconstruction.

4.1.3 Light source calibration w.r.t. the stereo-rig

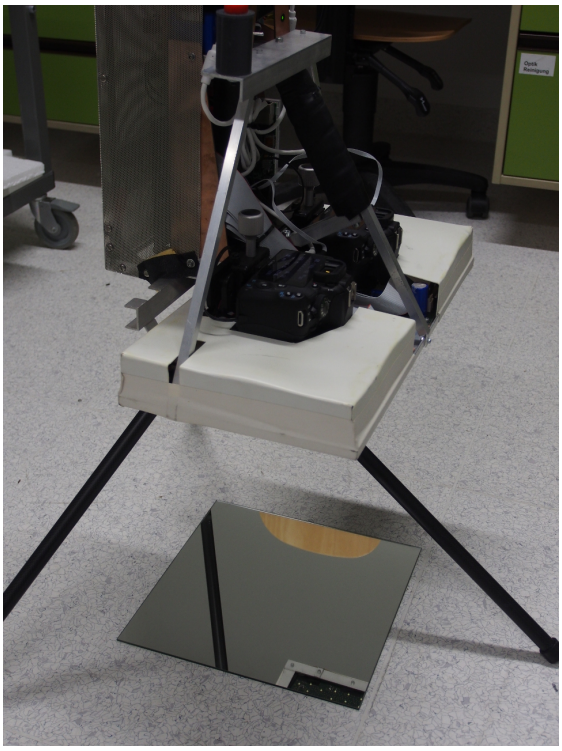
Finally, the light source consisting of 220 LEDs is calibrated. The LED board is composed of two distinct circuit-boards with 110 LEDs each. The position of the LEDs and their illumination direction is used to simulate the illumination, so it is essential to estimate these parameters. For the direction I assume that all LEDs of one circuit-board lie in the same plane. Using this assumption it suffices to estimate the normal of this plane which is the normal of the circuit-board. To obtain the relative position and orientation between the LEDs and the stereo-rig, I scan a mirror with reference markers, obtaining the 3D pose of the planar mirror relative to the stereo-rig, and capture an image pair with all LEDs switched on. Figure 4.5 shows the scanner with the reference target to estimate the mirror plane and an image pair of the mirrored LED board.

I extract LED positions from the camera images at subpixel accuracy (center of gravity of light blobs), match corresponding LEDs and reconstruct their position relative to the

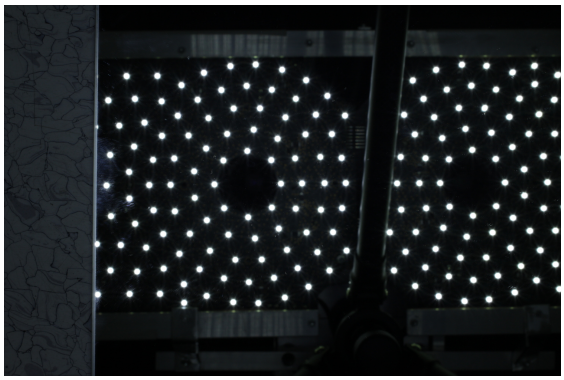
cameras using the theory of catadioptric vision [107]. To estimate the normal of the circuit board I use the position of the LEDs for one board to fit a plane. Figure 4.6 shows the result of the calibration and the sketch of the circuit-board design.



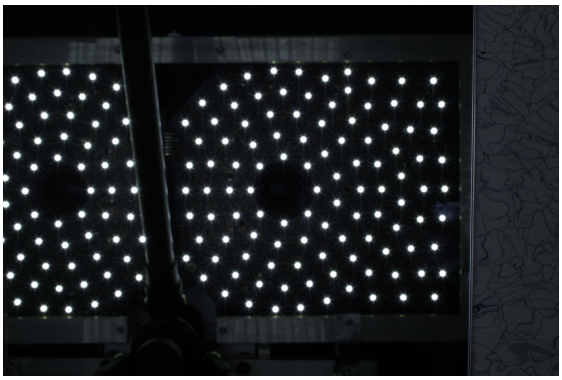
(a) Shows the scanner prototype with the reference target to estimate the pose of the mirror plane.



(b) Scanner and the planar mirror. A stereo pair of images is taken with the LEDs switched on. In (c) and (d) one can see a typical example of such a pair.

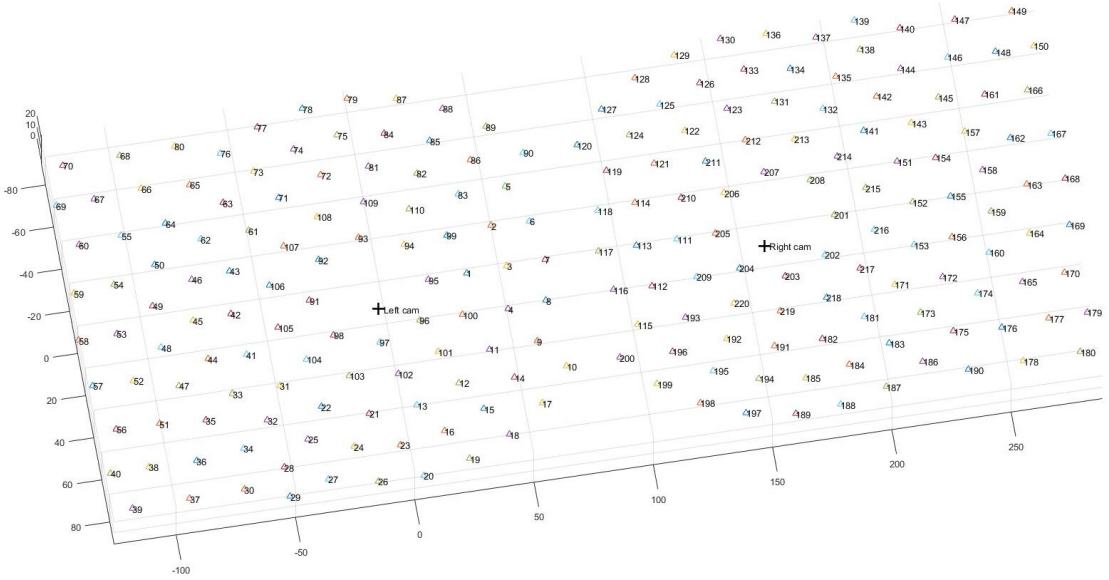


(c) Left image

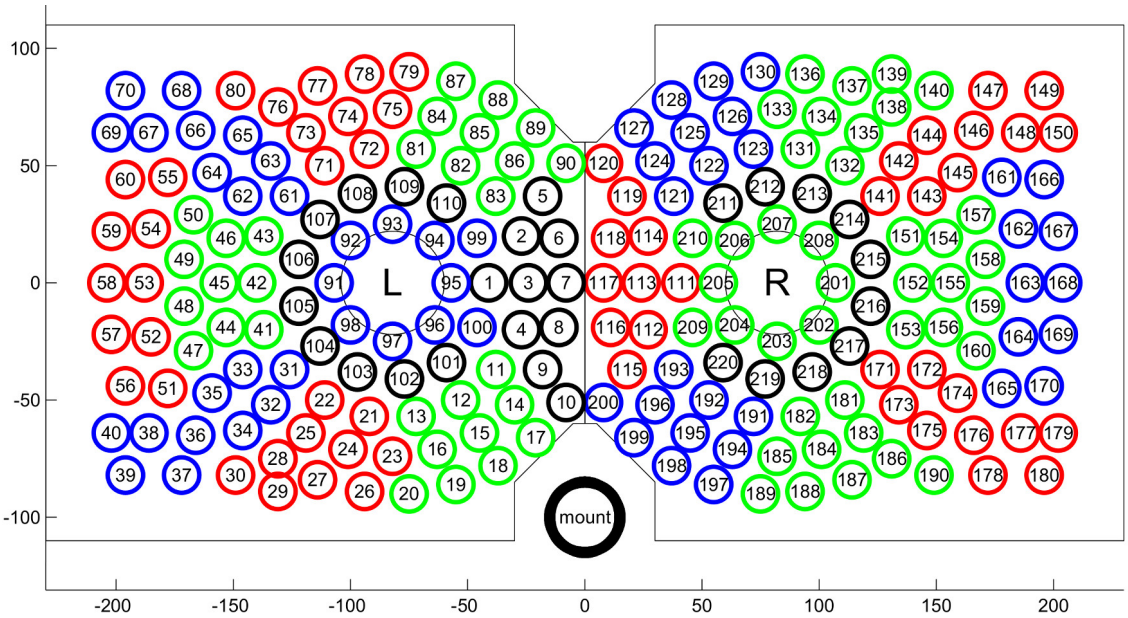


(d) Right image

Figure 4.5: This figure shows in (a) and (b) the scanner-prototype during LED board calibration and in (c) and (d) a stereo-image pair of the mirrored LED board.



(a) Result of the calibration: The positions for every LED and two surface normals (one for each circuit board).



(b) Sketch of the LED illumination showing the two circuit boards. In addition to control the intensity of all LEDs, groups of 10 LEDs each can be controlled individually as depicted by the differently colored circles.

Figure 4.6: Result of the LED board calibration and sketch of the two circuit-boards.

4.2 Estimating scene structure and camera poses

The 3D scene structure and the underlying camera poses are estimated in two steps. In the first step, the known calibration of the rigid triangular structure (left, right camera and prism, see fig. 4.2) is ignored. A first, rough estimate is calculated by applying Structure-from-Motion (SfM) to all individual camera images. Any available SfM algorithm might be used (e.g. [143],[57],[75]) at this stage, where all additional information about camera intrinsics and relative orientation of the stereo-rig and the 360° prism are ignored. In particular, in my experiments I used the SfM software of [75].
The result of this SfM processing is a similarity reconstruction of 3D surface points, individual camera poses, and estimated camera intrinsics. More formally, I obtain a set of K 3D points $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K\} \subseteq \mathbb{R}^3$, a set of J camera rotation matrices $\{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_J\} \subseteq \mathfrak{so}(3)$, a set of J camera centres $\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_J\} \subseteq \mathbb{R}^3$, and a set of J camera intrinsics.
In the second step, the knowledge about camera intrinsics and relative orientations (see fig. 4.2) is enforced in a *constrained bundle adjustment* process. As usual, 3D points and camera poses are iteratively refined, trying to minimize the reprojection error. In the presented approach, I fix the camera intrinsics according to the known calibration and constrain the camera poses w.r.t. the known relative orientation for all stereo pairs. Further, the georeferenced positions of the cameras are related to the tachymeter measurements via the calibrated vectors $\boldsymbol{t_l}$ and $\boldsymbol{t_r}$. I formulate this as a constrained non-linear optimization problem which is initialized with the SfM result from step 1 described above:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad f(\mathbf{x}) \quad \text{subject to} & \sum_{j}^{J/2} \boldsymbol{c}_j^{st}(\mathbf{x}) = \mathbf{0} \,, \\ & \sum_{j}^{M} \boldsymbol{c}_j^{ts}(\mathbf{x}, \mathbf{m}) = \mathbf{0} \,. \end{aligned} \tag{4.1}$$

In this optimization problem, I denote $\mathbf{x}$ as the set of variables $\mathbf{x} = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_K, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_J, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_J\}$ that are used to minimize the reprojection error $f(\mathbf{x})$, and $\mathbf{m} = \{\boldsymbol{m}_1 \ldots \boldsymbol{m}_M\}$ is a set containing M tachymeter measurements. I denote $\boldsymbol{c}_j^{st}$ as the constraints that enforce the relative stereo orientation between a camera pair and $\boldsymbol{c}_j^{ts}$ as the constraints that enforce the absolute position of a camera relative to the tachymeter measurements. To solve this optimization problem the *Ceres-solver* [12] within an *augmented Lagrangian multiplier* [48] scheme is used. Having defined the optimization problem in eq (4.1), I now describe the two constraints in the subsequent two paragraphs.

Stereo constraint: The stereo constraint ensures that the bundle adjustment finds a solution in which the relative orientation of a stereo camera pair is equal to the calibrated

relative orientation $\boldsymbol{R_s}, \boldsymbol{t_s}$. The stereo constraint for a camera j and its corresponding stereo camera j' is given by

$$\boldsymbol{c}_j^{st}(\mathbf{x}) = \boldsymbol{t_s} - \boldsymbol{R_s}\boldsymbol{R}_j\boldsymbol{c}_j + \boldsymbol{R}_{j'}\boldsymbol{c}_{j'} \pm \epsilon_b \ , \tag{4.2}$$

where $\boldsymbol{R_j} \in \mathbf{x}$ and $\boldsymbol{c_j} \in \mathbf{x}$ are the orientation and position of the camera j, $\boldsymbol{R_{j'}} \in \mathbf{x}$ and $\boldsymbol{c_{j'}} \in \mathbf{x}$ are the orientation and position of the corresponding stereo camera, and ϵ_b defines a tolerance interval reflecting small deviations during the calibration of the stereo-rig.

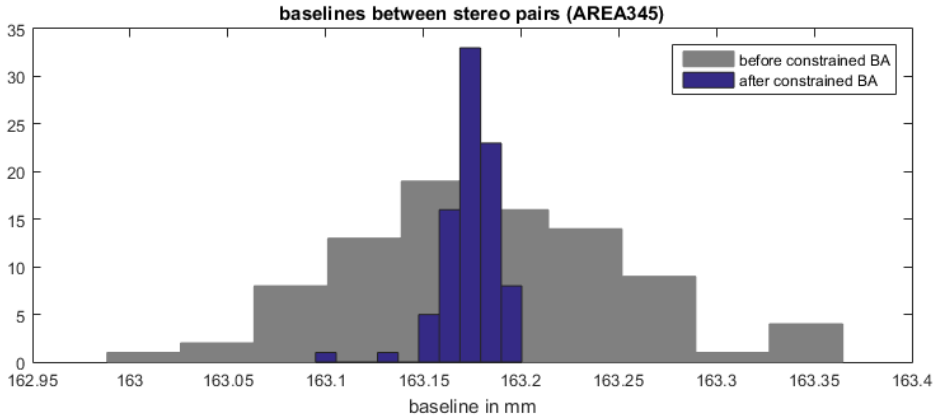


Figure 4.7: This figure demonstrates the decrease of the baseline deviation when applying the stereo constraint.

The effect of the stereo constraint can be seen in figure 4.7 where a histogram of baseline before and after applying the stereo constraint is shown. The first histogram (grey) displays the distribution of baseline values obtained after the SfM process, whereas the second one (blue) shows the distribution after the stereo constraint applied. One can see the decreased deviation after applying the stereo constraint.

Tachymeter measurement constraint: The tachymeter constraint enforces that the reconstructed prism position is equal to the position measured by the tachymeter. This also affects the camera pose because the prism and the camera are rigidly attached and their relative positions are known via the prism calibration. I define $\boldsymbol{m}_m$ as tachymter measurement that corresponds to camera j and define the constraint as follows:

$$\boldsymbol{c}_j^{ts}(\mathbf{x}, \mathbf{m}) = \begin{cases} \boldsymbol{R}_j(\boldsymbol{m}_m - \boldsymbol{c}_j) - \boldsymbol{t_l} \pm \epsilon, & \text{if left camera} \\ \boldsymbol{R}_j(\boldsymbol{m}_m - \boldsymbol{c}_j) - \boldsymbol{t_r} \pm \epsilon, & \text{if right camera} \end{cases} , \tag{4.3}$$

where $\boldsymbol{R}_j \in \mathbf{x}$ and $\boldsymbol{c}_j \in \mathbf{x}$ are the orientation and position of camera j, $\boldsymbol{m}_m \in \mathbf{m}$ is the corresponding tachymeter measurement, $\boldsymbol{t_l}$ and $\boldsymbol{t_r}$ are the relative positions of the prism w.r.t. left and right camera, and ϵ defines a tolerance interval reflecting the inherent imprecision of the tachymeter measurements.

In summary, by applying all the processing steps mentioned above, a georeferenced, Euclidean reconstruction of the 3D scene in terms of a sparse 3D point cloud, including all camera poses and all the individual positions and orientations of the LED light sources is obtained. In a final processing step, one can use SURE [132] (or other methods like PMVS [53]), to densify the 3D point cloud so that the final result of the geometric processing consists of:
(a) J Camera poses and their corresponding images,
(b) for every camera pose, N positions and orientations of the LEDs, and
(c) a dense 3D point cloud representing the 3D surface geometry.
This constitutes the required geometric knowledge to perform "radiometry in the wild" as described in chapter 5.

The constrained bundle adjustment presented here is an iterative method. The iteration terminates, if a maximum number of iterations (default is 50 iterations) is exceeded or when all constraints fulfil a certain constraint tolerance (the default value is 0.1mm).
I choose ϵ_b and ϵ in such a way that they are in the order of magnitude of the of the expected calibration (sub-mm range) and tachymeter measurement (mm range) tolerance. For the baseline constraint, as default value for ϵ_b, I used $\epsilon_b = 0.1$mm. This means that the difference between the calibrated baseline and the estimated baselines must not exceed 0.1mm. For the tachymeter measurement constraint, I used as a default value, $\epsilon = 2.0$mm.

4.3 Experiments

To validate the presented approach I conduct a series of experiments that quantify the performance of (a) the quality of the geo-referencing and (b) the quality of the surface reconstruction w.r.t. ground truth measurements.

4.3.1 Dataset: AREA345

Figure 4.8a shows a dense 3D point cloud of *Area345* on *Seradina 12C* (see section 3.3). The area shown here was scanned from 87 individual scanner positions including

87 tachymeter measurements[2], resulting in a total of 348 images, 174 with and 174 without LED illumination (more details on using with-and-without-LED image pairs can be found in chapter 5). The diameter of this scene is about 940mm and it took 45 minutes to scan the scene. I use tachymeter measurements of artificial reference markers (green circles) to evaluate the georeferenced accuracy. For the area indicated by the red rectangle, I use a dense 3D point cloud obtained by a structured-light scan as ground truth data (see section 3.3.3).

4.3.1.1 Quality of the geo-referencing

In a separate tachymeter measurement, I measure the coordinates of the four reference markers (green circles in fig. 4.8a) and compare their 3D position with the reconstructed position in the 3D point cloud. The mean distance between our reconstruction and the tachymeter measurements is 12.4mm which is in the typical range of the accuracy of a tachymeter measurement.

4.3.1.2 Comparison with ground truth data

Figure 4.8 shows the 3D reconstruction of a smaller area (fig. 4.8b), and the distance between our reconstruction and ground truth (fig. 4.8c). The root-mean squared distance is 0.13mm.

4.3.2 Dataset: Ser12c

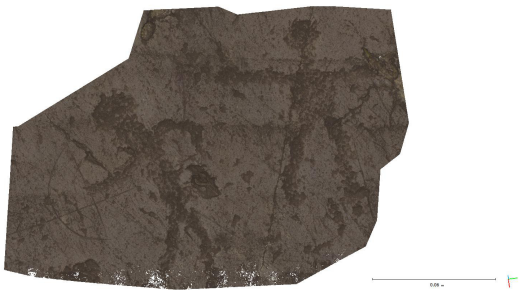
This dataset is the result of two different scanning methods. Figure 4.9a and figure 4.9b show the final geo-referenced reconstruction of the rock panel *Seradina I rock 12 C* obtained using a micro aerial vehicle (see section 3.3.1) with superimposed geo-referenced reconstructions obtained by the presented scanning approach (marked with the two yellow bounding boxes) (see section 3.3.3). In figure 4.9c one can see a close-up of the reconstruction and in pseudo-color the absolute distance between the reconstruction of the rock panel and the reconstructions obtained by the scanner prototype (the histograms in 4.9d show the distance distribution of the two reconstructions).
As the histograms in figure 4.9d suggest, the mean absolute distance between the reconstruction obtained by the MAV and our scanner is about $0.9 - 1.1$cm.

4.3.3 Dataset: Camunian rose

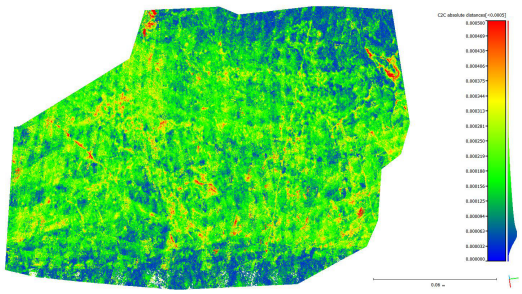
In this experiment we validate the scanners reconstruction against the ground truth of two high precision 3D prints. Figure 4.10 shows a pseudo-color visualization of the absolute

[2]Leica TS11 1" R500.

(a) Entire AREA345 reconstruction



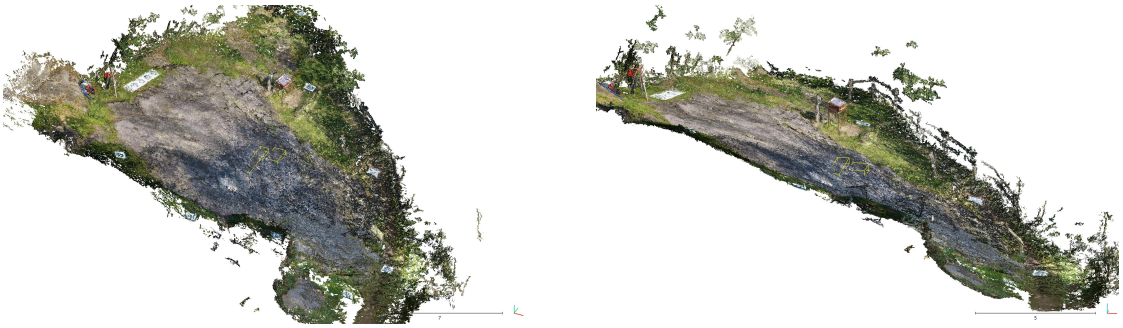
(b) Dense, radiometrically calibrated 3D point cloud



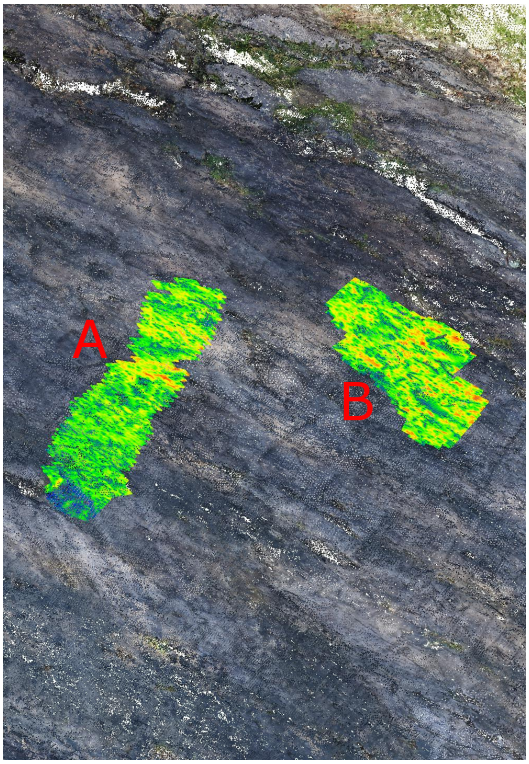
(c) Pseudo-color visualization of distances w.r.t. ground truth

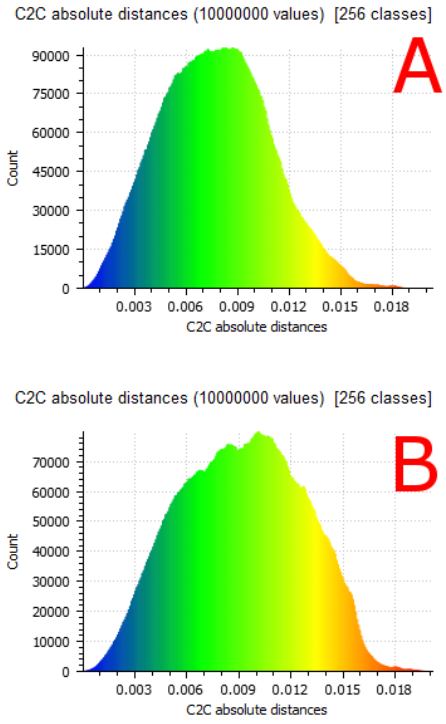
Figure 4.8: In (a) one can see the dense, radiometrically calibrated 3D point cloud of Area345 on Seradina 12C with prehistoric rock-art. The four artificial markers circled in green have been placed in the scene to validate the quality of the geo-referencing. The red rectangle highlights the area used for validation w.r.t. ground truth (see fig. 4.8c).
In (b) and (c) one can see a zoom into the area marked in red in fig. 4.8a. (b) the 3D reconstruction, and (c) the pseudo-color visualisation of the absolute distance between the reconstructed point cloud and the ground truth data (obtained by independent structured-light scans). Distances larger than 0.5mm are clipped. The root-mean squared distance between ground truth and reconstruction is 0.13mm.

distance between the 3D print and the obtained dense surface reconstruction. The mean absolute distance between the reconstruction and the 3D print is 0.11 mm.



(a) (b)





(c) (d)

Figure 4.9: Reconstruction obtained by two different scanning methods. (a) and (b) show the fusion of two geo-referenced reconstructions (Figure 4.1 shows a close up of (a)). First, a large rock panel (Seradina 12C) obtained by a micro aerial vehicle and second, highlighted by the yellow bounding boxes, the reconstructions obtained by the scanner prototype and the presented approach. (c) and (d) state the absolute distance between the rock panel reconstruction and the reconstruction obtained by the scanner prototype (units in meter). "A" corresponds to Area10, "B" to Area345 (see section 3.3.3).

4.3.4 Processing times

The time required for the full dense reconstruction is dominated by the *densification* of the sparse SfM reconstruction. In this work I use SURE [132] to accomplish this.

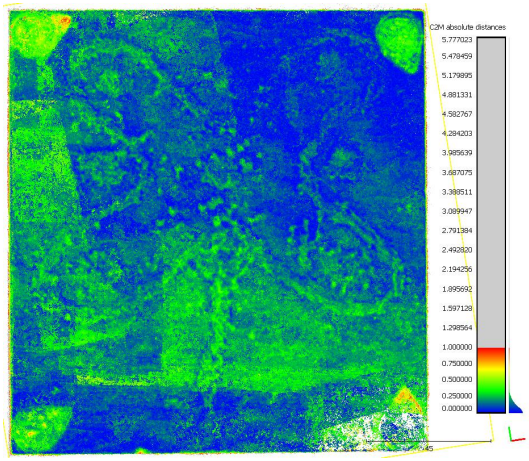


Figure 4.10: This image shows a pseudo-color visualization of the distance between the 3D print and the reconstruction.

The used SfM pipeline (based on [75]) performs the following steps: it un-distorts the images, extracts features, matches the features, estimates and verifies the epipolar geometry, reconstructs camera poses and sparse scene points (including incremental bundle adjustment), and solves the final constrained bundle adjustment optimization.
Table 4.3 states typical processing times for different numbers of images.

Number of images	Approx. number of features per image	Time (sparse)	Time (dense)
86	2500	4 min	4 hours
1052	2500	4 hours	14 hours

Table 4.3: This table states typical processing times for the geometric reconstruction. The total time for a full dense reconstruction is dominated by the *densification* step. The resulting time for the dense point cloud with 1052 images is obtained with a reduced image resolution of $\frac{1}{2}$ of the original image size.

4.4 Discussion

This chapter introduced the geometric reconstruction method, used in the 3D Pitoti project for micro-scale reconstruction. Using this method one is able to produce a Euclidean reconstruction of the scene without the need for a further scale estimation step. In addition, if the scanner prototype is used in conjunction with a tachymeter, the presented method is capable to produce a geo-referenced Euclidean reconstruction.
During the project and the thesis we could identify additional benefits and potential improvements:

The scanner can be operated by non-experts in the field of photogrammetry: Over the project duration, archaeologists used the scanner prototype during their field work several times and where able to scan and reconstruct the rock-art. This has mainly two reasons: first, the on-site feedback provided by the online structure from motion software, and second, the scanning principle. Due to the mini-tripod of the scanner the orientation of the image plane w.r.t. the rock surface is almost parallel. The distance between scanner and surface is almost constant. Because of this arrangement, the scanning is done in a similar way as the scanning in aerial photography. The users, by default, rasterize the surface in a systematic way so that they prevent holes in the scan and because of the constant distance and orientation of the cameras almost all images have the same high spatial resolution.

The fusion of different geometric constraints is challenging: The presented approach fuses three different geometric constraints: first, the scanner calibration which includes the cameras, stereo-rig, and the position of the 360° prism w.r.t. the cameras. Second, the scene structure and camera poses, estimated by structure from motion, and third, the tachymeter measurements.
In the presented approach I fuse these three constraints using a constrained bundle adjustment procedure, which is a non-linear constrained minimization problem (as stated in section 4.2). Each of these constraints has its own uncertainty about the "true" measurement (For example, the uncertainty of the tachymeter measurement with the mounted prism is about $2-4mm$ [81] and depends also on the skills of the operator). These differences can lead to a possibly inconsistent minimization problem with the effect of distorting the final geometric reconstruction. To tackle this issue, I perform the minimization in the following way:

1. I use an *augmented Lagrangian multiplier* (ALM) method for the minimization. This method belongs to the *penalty methods* in the field of constrained minimization [48]. This group of algorithms penalize a minimization solution, if some constraints are not satisfied, by increasing the importance (weights) of those constraints. In my first paper [73] on this topic, I used the *quadratic penalty* method in which I have to set the weight manually which is hard to do in advance and can lead to a highly distorted cost function [48]. Hence, for my second paper [74], I decided to use ALM because this method determines the weight automatically and estimates an offset to overcome the problem with the distorted cost function. For more details on this topic I refer the reader to the book of R. Fletcher [48].
2. I define a *tolerance box* in which the constraints can vary without affecting the cost

function. In equation 4.2 and equation 4.3, ϵ_b and ϵ reflect the different uncertainties of the stereo-rig calibration and the tachymeter measurements. In this way I can handle different magnitudes of uncertainties.

A more mathematical profound way would be to model the uncertainty with probabilities and incorporate this knowledge in the minimization.

Despite the challenge of fusing different constraints, the benefit of the presented method is that it allows to include more information to solve the non-linear optimization. Minimizing a non-linear problem often leads to local minima solutions. Using *e.g.* the presented *stereo constraint* (equation 4.2) one can avoid these local minima in which the stereo constraint is violated.

Drift compensation: In the case where the scanner is used in conjunction with a tachymeter, the presented approach compensates drift induced by the SfM reconstruction. The same behaviour cold be achieved, if geo-referenced *ground control points* would be used (which was not desired in the 3D Pitoti project).

Using the LED illumination to estimate surface normals: In the current approach the reconstructed geometry is a point cloud. Surface normals, which will play an important role in the next chapter, are estimated based on the acquired point cloud, assuming that the resolution is dense enough to do so (the highest resolution of the resulting point cloud is in the range of $440k$ points per cm^2).

With our custom built LED-illumination I have the opportunity to flash individual groups of LEDs (see *e.g.* figure 4.6). By doing so I could record a set of images of the surface under varying incident illumination. Using this set of images one could perform a photometric stereo [165] reconstruction which includes the surface normals.

5

Radiometric reconstruction

One of the challenges in scanning radiometric surface properties “in the wild” is the illumination condition at the rock-art sites. Different from the laboratory one is confronted with non-controllable illumination conditions, direct sunlight, and cast shadows caused for example by nearby trees. Despite of these conditions the proposed approach is able to deliver radiometric surface properties beyond photo-texture without the need of shrouding or color reference targets.

In a first processing step the influence of ambient illumination is eliminated by taking pairs of images with and without the LED-flash of the scanner. Subtracting these two frames from each other results in images as if taken flash-only in complete darkness, under perfect shrouding of daylight. In a second step the knowledge about (a) camera pose, (b) illumination direction and intensity, and (c) 3D surface geometry is combined to obtain

precise measurements of radiometric surface properties at each surface point.
The final result of the combined, geometric and radiometric processing is a dense 3D point cloud with radiometric surface properties mapped to each of the 3D points or a set of re-rendered images with the radiometric surface properties mapped to each pixel.

5.1 Theory, terminology and notation

I start to establish the terminology and notation used in the remainder of this chapter by reviewing theoretical background, first describing the observed surface radiance, followed by explaining how this radiance is mapped to a camera pixel value. I use the *reflectance equation* [117] to model the surface radiance

$$L_o(\boldsymbol{p}, \boldsymbol{\omega_o}) = \int_{\mathcal{H}^2(\boldsymbol{n_p})} f(\boldsymbol{p}, \boldsymbol{\omega_o}, \boldsymbol{\omega_i}) L_i(\boldsymbol{p}, \boldsymbol{\omega_i}) \cos(\theta_i) d\boldsymbol{\omega_i} \ , \tag{5.1}$$

where L_o denotes the radiance of surface point $\boldsymbol{p}$ reflected in direction $\boldsymbol{\omega_o}$. $L_i(\boldsymbol{p}, \boldsymbol{\omega_i})$ denotes the incident radiance from direction $\boldsymbol{\omega_i}$, and $f(\boldsymbol{p}, \boldsymbol{\omega_o}, \boldsymbol{\omega_i})$ is the *bidirectional reflectance distribution function* (BRDF) [113]. Equation (5.1) models the observed radiance L_o at a surface point $\boldsymbol{p}$ as the integral of the incident radiance L_i weighted by the BRDF at this point and by the falloff term $\cos(\theta_i)$. Here, θ_i is the angle between the surface normal $\boldsymbol{n_p}$ and the incident radiance direction $\boldsymbol{\omega_i}$ (it follows that $\cos(\theta_i) = \boldsymbol{n_p} \cdot \boldsymbol{\omega_i}$). I use the subscript $\boldsymbol{p}$ for the normal vector $\boldsymbol{n_p}$ to explicitly state that the normal vector is evaluated at surface point $\boldsymbol{p}$. In this general formulation, the integral is calculated over the upper hemisphere $\mathcal{H}^2(\boldsymbol{n_p})$.
The fact that our scanner prototype uses a discrete set of N directed light sources, replaces the integral in eq (5.1) by a finite summation. Hence eq (5.1) can be rewritten as

$$L_o(\boldsymbol{p}, \boldsymbol{\omega_o}) = \sum_i^N f(\boldsymbol{p}, \boldsymbol{\omega_o}, \boldsymbol{\omega_i})(\boldsymbol{n_p} \cdot \boldsymbol{\omega_i}) \frac{I_i(\boldsymbol{\omega_i} \cdot \boldsymbol{\omega_{I_i}})}{r_i^2} \ , \tag{5.2}$$

where $I_i(\boldsymbol{\omega_i} \cdot \boldsymbol{\omega_{I_i}})$ represents the radiant intensity of light source i as a function of the direction of the light source $\boldsymbol{\omega_{I_i}}$ and the incident direction $\boldsymbol{\omega_i}$ at surface point $\boldsymbol{p}$, with r_i describing the distance between surface point and light source.
For the case of Lambertian reflectance, $f(\boldsymbol{p}, \boldsymbol{\omega_o}, \boldsymbol{\omega_i}) = \frac{\rho(\boldsymbol{p})}{\pi}$, so that one obtains

$$L_o(\boldsymbol{p}, \boldsymbol{\omega_o}) = \frac{\rho(\boldsymbol{p})}{\pi} \sum_i^N (\boldsymbol{n_p} \cdot \boldsymbol{\omega_i}) \frac{I_i(\boldsymbol{\omega_i} \cdot \boldsymbol{\omega_{I_i}})}{r_i^2} \ , \tag{5.3}$$

where $\rho(\boldsymbol{p})$ is the Lambertian reflectance property at $\boldsymbol{p}$. Figure 5.1 illustrates the imaging configuration corresponding to eq 5.3.

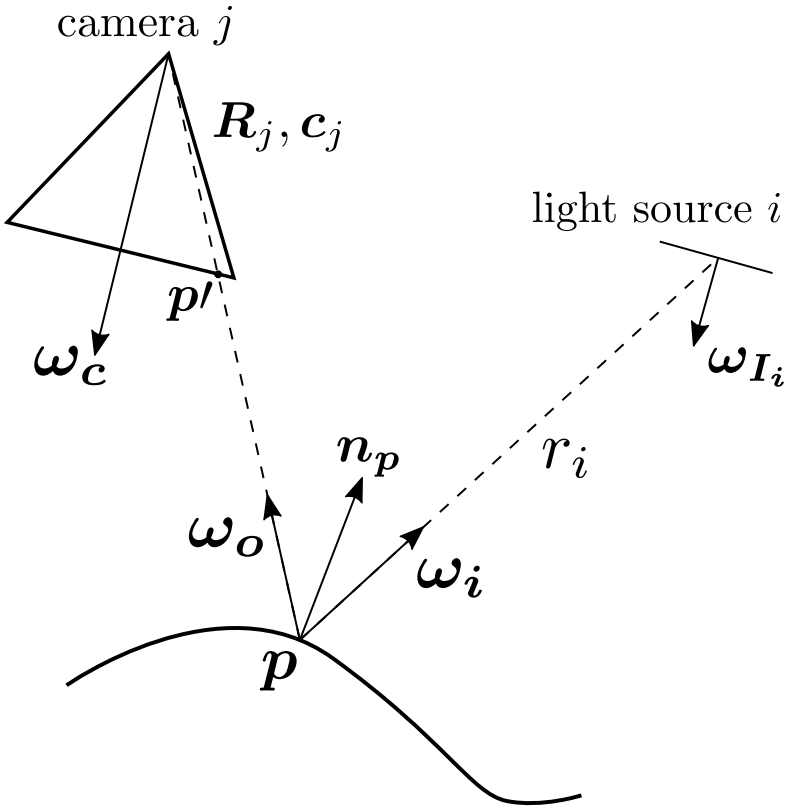


Figure 5.1: Imaging configuration corresponding to eq 5.3, showing one camera position j and one directed light source i with respect to surface point $\boldsymbol{p}$ and surface normal $\boldsymbol{n_p}$.

To further emphasize, which parts of $L_o(\boldsymbol{p}, \boldsymbol{\omega_o})$ depend on the *geometric configuration* defined by scene (surface normal) and illumination, I let geometry

$$G(\boldsymbol{n_p}, \boldsymbol{\omega_1}, \ldots, \boldsymbol{\omega_N}, \boldsymbol{\omega_{I_1}}, \ldots, \boldsymbol{\omega_{I_N}}) = \frac{1}{\pi} \sum_{i}^{N} (\boldsymbol{n_p} \cdot \boldsymbol{\omega_i}) \frac{I_i(\boldsymbol{\omega_i} \cdot \boldsymbol{\omega_{I_i}})}{r_i^2} \,. \tag{5.4}$$

Equation (5.3) can now be rewritten in abbreviated notation:

$$L_o(\boldsymbol{p}, \boldsymbol{\omega_o}) = \rho(\boldsymbol{p}) G \,. \tag{5.5}$$

So far, I have presented how I model the interaction between surface and light. The radiance reflected by the surface in direction $\boldsymbol{\omega_o}$ is now captured by a camera. The relationship between the resulting irradiance on the camera sensor $E(\boldsymbol{p\prime})$ and the reflected radiance $L_o(\boldsymbol{p}, \boldsymbol{\omega_o})$ is given by

$$E(\boldsymbol{p\prime}) = V(\boldsymbol{\omega_c}, \boldsymbol{\omega_o}) L_o(\boldsymbol{p}, \boldsymbol{\omega_o}) \,, \tag{5.6}$$

where $\boldsymbol{p\prime}$ is the pixel corresponding to surface point $\boldsymbol{p}$, and $V(\boldsymbol{\omega_c}, \boldsymbol{\omega_o})$ models the vignetting of real world camera lenses by

$$V(\boldsymbol{\omega_c}, \boldsymbol{\omega_o}) = \frac{\pi}{4} \left(\frac{d}{f} \right)^2 (\boldsymbol{\omega_c} \cdot \boldsymbol{\omega_o})^4 \,. \tag{5.7}$$

Here, d is the diameter of the camera lens, f is the focal length of the lens, and $\boldsymbol{\omega_c}$ is the direction of the optical axis of the camera [77]. We further assume a linear camera response so that the measured pixel value $P(\boldsymbol{p\prime})$ is proportional to the irradiance on the camera sensor. Using equations (5.5) and (5.6) to reformulate $E(\boldsymbol{p\prime})$, we obtain the following relation between surface reflectance $\rho(\boldsymbol{p})$, camera vignetting V and geometric configuration G:

$$P(\boldsymbol{p\prime}) \propto E(\boldsymbol{p\prime}) = \rho(\boldsymbol{p}) V G \,. \tag{5.8}$$

The estimation of the proportional relationship between $P(\boldsymbol{p\prime})$ and $E(\boldsymbol{p\prime})$ can be seen as a radiometric calibration of the camera-illumination setup and is discussed in further detail in section 5.2.

In the presented approach I am interested in the Lambertian reflectance term $\rho(\boldsymbol{p})$ which can be computed if the geometric configuration G, the vignetting term V, and the proportional relationship between $P(\boldsymbol{p\prime})$ and $E(\boldsymbol{p\prime})$ are known. The results of chapter 4 are used to calculate G, and in section 5.2 I show how to estimate the vignetting term V and the proportional relationship $P(\boldsymbol{p\prime}) \propto E(\boldsymbol{p\prime})$.

5.2 Radiometry in the wild

The goal of the radiometric processing described in this section is twofold: first, I aim at estimating the surface reflectance properties $\rho(\boldsymbol{p})$ of each 3D point $\boldsymbol{p}$ in the scene based on all camera images taking the particular illumination situation for each image into account; second, I wish to map these surface properties to the corresponding pixels in all cameras, rendering new images $\rho(\boldsymbol{p\prime})$ at pixels $\boldsymbol{p\prime}$. Thus the result of "radiometry in the wild" is a new set of J images with their pixel values corresponding to the observed surface reflectance. This processing departs from eq 5.8, $P(\boldsymbol{p\prime}) \propto \rho(\boldsymbol{p})VG$, and is performed in four distinct processing steps as outlined below.

I Frame differencing In the first radiometric processing step, the influence of daylight or any other ambient illumination is eliminated by frame differencing. The scanner prototype takes two *linear RAW* images in rapid succession, one image with LED illumination switched on, the other one without artificial illumination. Frame differencing yields

$$P_{FD}(\boldsymbol{p\prime}) = P_{LED}(\boldsymbol{p\prime}) - P_{dark}(\boldsymbol{p\prime}) \,, \tag{5.9}$$

with $P_{LED}(\boldsymbol{p\prime})$ the pixel value including, and $P_{dark}(\boldsymbol{p\prime})$ the corresponding pixel value without LED illumination.
Figure 5.2 shows the camera response of a linear RAW image of the grey values on the color reference target (used camera: Canon EOS 100D). A reconstruction of the color reference target can be seen in figure 5.5a.

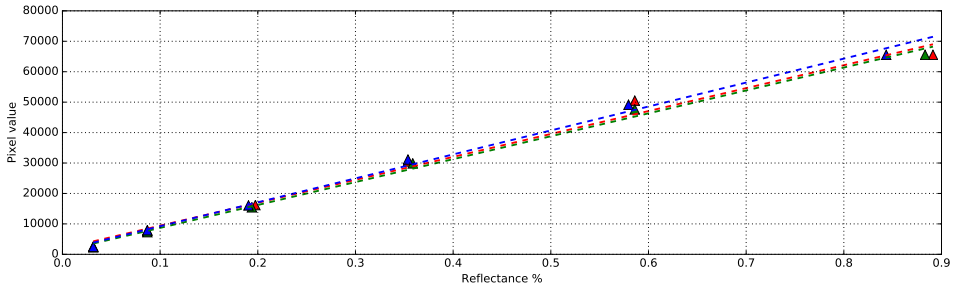


Figure 5.2: Camera response of the three RGB channels of the used Canon EOS 100D camera. The pixel values of the grey patches of a color reference target are shown against their measured reflectance. The lines are fitted using the acquired data points.

II Vignetting compensation In the next step, the effect of vignetting is compensated. The vignetting term V (eq 5.7) can be easily calculated because of the known geometry

and camera intrinsics. Hence, I obtain

$$P_{FD,V}(\boldsymbol{p\prime}) = \frac{P_{FD}(\boldsymbol{p\prime})}{V} \propto \rho(\boldsymbol{p})G \,, \tag{5.10}$$

where $P_{FD,V}(\boldsymbol{p\prime})$ denotes the pixel value after frame differencing and vignetting compensation.

III Simulating homogeneous surface illumination The incident light depends on the geometric configuration G. For each camera, there exist 220 LEDs that illuminate the scene. The amount of light that reaches a point on the surface depends on the radiant intensity of the LEDs, the direction of the light sources, the distance between the light sources and the surface point, and the surface normal.

The LEDs show a radiant intensity characteristic $I_i(\boldsymbol{\omega_i} \cdot \boldsymbol{\omega_{I_i}})$ that depends on the radiation angle $\arccos(\boldsymbol{\omega_i} \cdot \boldsymbol{\omega_{I_i}})$ and is provided by the LED manufacturer. I use this specification to model the LED radiant intensity. The positions and orientations of the LEDs w.r.t. the scene are obtained by the geometric calibration and reconstruction processes described in chapter 4.

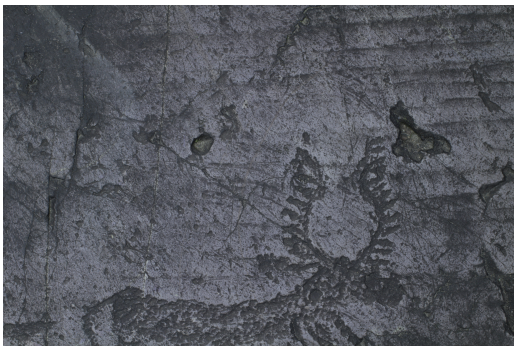
For the surface geometry representation, I build on the dense 3D point cloud obtained by SURE. In a camera centric representation, it makes sense to compute the depth-map of a part of the scene, as observed by the camera. For this, I project the dense 3D surface points on the image plane and store the corresponding depth in a per-pixel depth-map. In case of small holes in the depth-map, these are filled by nearest neighbour approximation. Next, I use this depth-map to compute surface normals and the distances between the surface points and the LEDs. Figure 5.3 shows these processing steps.

I am now able to calculate G and can further rewrite equation 5.10 as

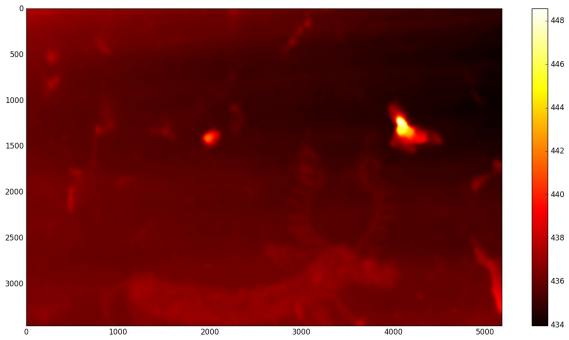
$$\rho(\boldsymbol{p}) \propto \tilde{P}(\boldsymbol{p\prime}) = \frac{P_{FD,V}(\boldsymbol{p\prime})}{G} \,, \tag{5.11}$$

where $\tilde{P}(\boldsymbol{p\prime})$ represents the intensity of a particular pixel under the assumption that the whole surface is homogeneously illuminated.

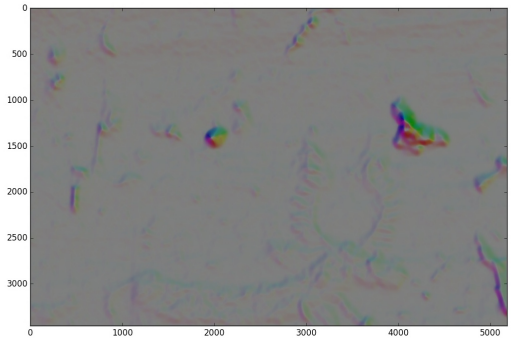
IV Radiometric calibration Up to this point, I have modeled pixel intensities and surface radiance under the assumption of homogeneous illumination, but completely neglected spectral reflectance properites of the surface and spectral characteristics of the sensor. In fact, I obtain three color measurements $\tilde{P}_r(\boldsymbol{p\prime}), \tilde{P}_g(\boldsymbol{p\prime}), \tilde{P}_b(\boldsymbol{p\prime})$ for each pixel which must be related to the actual scene reflectance $\rho_r(\boldsymbol{p}), \rho_g(\boldsymbol{p}), \rho_b(\boldsymbol{p})$ in red, green and



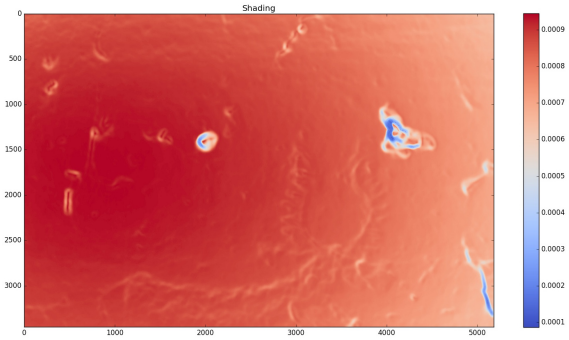
(a) Original image



(b) Estimated depth map



(c) Estimated normal map



(d) Estimated incident illumination G

Figure 5.3: This figure shows exemplarily the estimation of the incident illumination. (a) the original image. Because of the preceding dense surface reconstruction I can calculate the corresponding depth (b) and normal map (c) of the image. Finally, the incident illumination G is calculated (d).

blue. This is achieved by an affine mapping $\boldsymbol{M}$ from $\tilde{P}$ to ρ

$$\begin{pmatrix} \rho_r(\boldsymbol{p}) \\ \rho_g(\boldsymbol{p}) \\ \rho_b(\boldsymbol{p}) \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} \tilde{P}_r(\boldsymbol{p\prime}) \\ \tilde{P}_g(\boldsymbol{p\prime}) \\ \tilde{P}_b(\boldsymbol{p\prime}) \\ 1.0 \end{pmatrix}, \tag{5.12}$$

where $\boldsymbol{M} \in \mathbb{R}^{3\times4}$ is a 3×4 matrix that encodes the proportionality between scene radiance and pixel value in eq 5.11. Therefore, eq 5.12 directly relates the spectral reflectance of a surface point $\boldsymbol{p}$ to the corresponding color measurement at pixel $\boldsymbol{p\prime}$.

In the final step of radiometric processing, I estimate the matrix $\boldsymbol{M}$ by scanning a 3D scene that contains a color reference target with known spectral reflectance values for each of its S color patches [1]. This leads to S correspondences between known reflectance values of

[1] I used *CIE RGB* color values [169]: gamma: 2.2; reference white: *CIE illuminant E*; Primaries: (xr,

the target $(\rho_{r,s}, \rho_{g,s}, \rho_{b,s})^T$ and average measured pixel values $(\tilde{P}_{r,s}, \tilde{P}_{g,s}, \tilde{P}_{b,s})^T, s = 1 \dots S$. Now, I follow the algorithm of [164] and estimate $\boldsymbol{M}$ by solving this overdetermined system of linear equations.

After radiometric calibration, knowing $\boldsymbol{M}$, I map surface points to pixels and render new images obtained by substitution $\boldsymbol{p} \rightarrow \boldsymbol{p\prime}$:

$$P_\rho(\boldsymbol{p\prime}) = \begin{pmatrix} \rho_r(\boldsymbol{p\prime}) \\ \rho_g(\boldsymbol{p\prime}) \\ \rho_b(\boldsymbol{p\prime}) \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} \tilde{P}_r(\boldsymbol{p\prime}) \\ \tilde{P}_g(\boldsymbol{p\prime}) \\ \tilde{P}_b(\boldsymbol{p\prime}) \\ 1.0 \end{pmatrix} . \tag{5.13}$$

This pixel-level computation is performed for all pixels in all of the J images. The final result of radiometric processing is a set of J new, radiometrically calibrated images with direct relation between pixel color and measured spectral surface reflectance. Figure 5.4 illustrates all the individual processing steps described above, for the example of a color reference target scanned in direct sunlight, including a cast shadow.

Finally, results from chapter 4 and 5.2 can be combined to produce dense, radiometrically calibrated 3D point clouds as shown in the various examples provided below.

5.3 Experiments

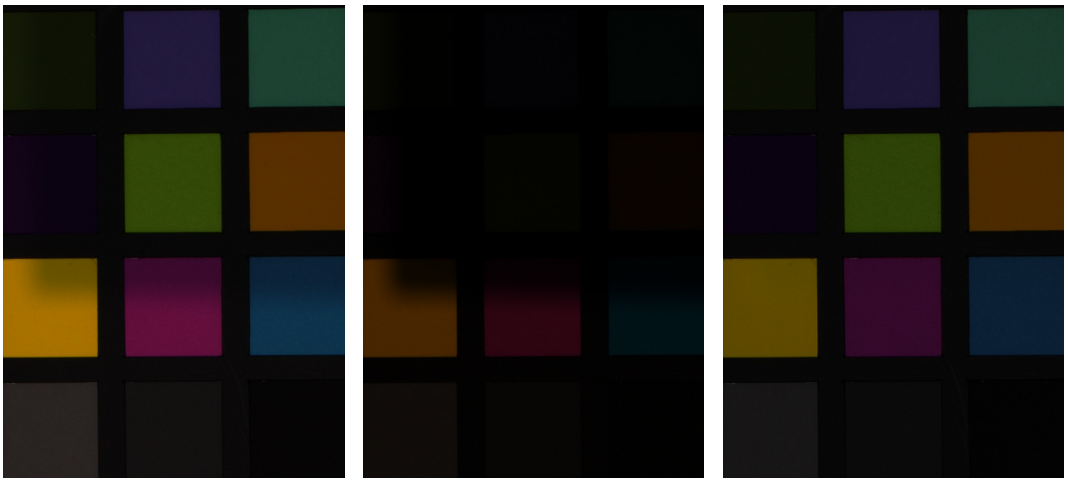
5.3.1 Radiometric validation w.r.t. ground truth

I validate the radiometric reconstruction approach by comparing the results with the known radiometric characteristics of a color reference target, using three different test scenes of varying complexity as shown in fig. 5.5:

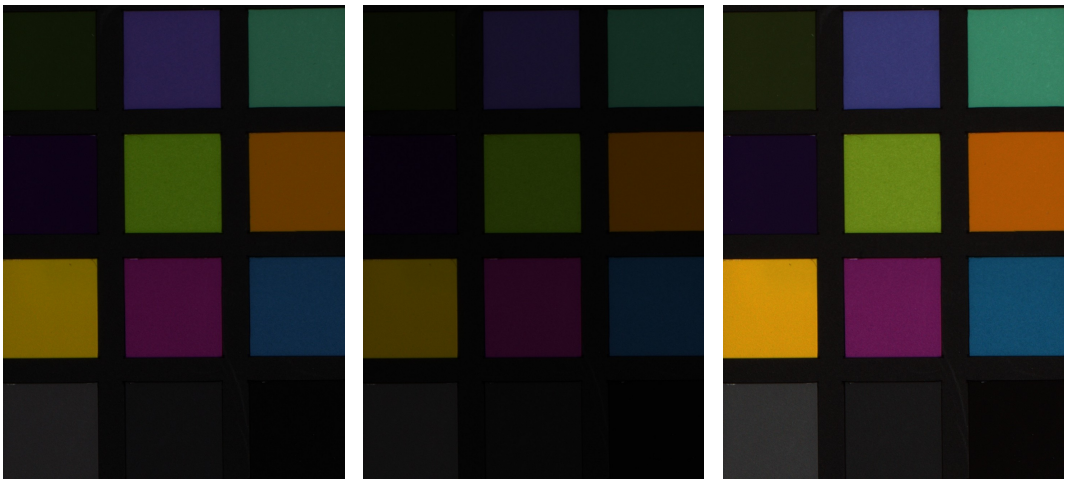
Calibration scene: This scene consists of the color target alone and was captured in a completely dark room. The color target was placed on the floor so that the planes of the LED illumination and the color target were approximately parallel. 13 scanner positions were used to acquire this scene.

Daylight scene: This scene consists of the color target placed on a wooden floor, captured under bright daylight, including direct sunlight and cast shadows. Again, target and LED planes are approximately parallel. For this scene I used 7 scanner positions. See

yr, Y) = (0.7350, 0.2650, 0.176204), (xg, yg, Y) = (0.2740, 0.7170, 0.812985), (xb, yb, Y) = (0.1670, 0.0090, 0.010811)



(a) $P_{LED}(\boldsymbol{p\prime})$ (b) $P_{dark}(\boldsymbol{p\prime})$ (c) $P_{FD}(\boldsymbol{p\prime})$



(d) $P_{FD,V}(\boldsymbol{p\prime})$ (e) $\tilde{P}(\boldsymbol{p\prime})$ (f) $P_{\rho}(\boldsymbol{p\prime})$

Figure 5.4: This figure illustrates all the individual radiometric processing steps, from image capture with and without LED illumination up to the final radiometrically corrected image P_{ρ}. The images show part of a color reference target captured in direct sunlight, including a shadow cast by the scanner (see fig. 5.5b for the complete reconstruction result of this scene).

fig. 5.4 for detail views of images from one of the scanner positions.

Evaluation scene: In this scene, I placed the color target such that the planes of the target and the LED illumination are not parallel, so that the illumination density differs from the previous two settings. Further, the scene contains a piece of rock with a rough 3D

(a) Calibration scene

(b) Daylight scene

(c) Evaluation scene

Figure 5.5: Dense, radiometrically calibrated 3D point clouds of the three scenes used for radiometric validation.

surface that represents the usual operational scenario of the scanner. 20 scanner positions were used to acquire this scene.

I use these three scenes for a number of detailed quantitative experimental validations described below. Color distances are given in two different metrics in *Lab* color space: Euclidean distance ΔE^*_{ab}, and $CIEDE2000$ color difference metric ΔE_{CIE00} [128] which is claimed to better model human perception of color differences.

	#img	$\mu(\Delta E^*_{ab})$	$\sigma(\Delta E^*_{ab})$	$\mu(\Delta E_{\text{CIE00}})$	$\sigma(\Delta E_{\text{CIE00}})$	$\Delta E^*_{ab,GT}$	$\Delta E_{\text{CIE00, GT}}$
dark skin	17	1.42	0.62	1.08	0.49	3.82	2.50
light skin	17	1.13	0.50	0.78	0.37	5.81	3.05
blue sky	17	1.64	0.72	1.52	0.72	2.37	2.22
foliage	18	1.58	0.74	0.99	0.51	2.68	1.98
blue flower	19	1.40	0.52	1.05	0.44	1.39	0.78
bluish green	18	1.63	0.83	0.81	0.34	2.47	1.44
orange	17	1.47	0.74	0.65	0.36	12.37	3.07
purplish blue	17	1.39	0.67	0.94	0.50	4.11	1.69
moderate red	17	1.26	0.59	0.65	0.27	1.83	0.95
purple	18	1.24	0.60	0.72	0.37	5.52	2.84
yellow green	18	1.74	0.74	0.75	0.28	4.24	1.56
orange yellow	18	1.45	0.78	0.86	0.45	7.82	1.96
blue	14	1.79	0.93	0.91	0.44	7.24	2.39
green	14	1.45	0.54	0.76	0.29	3.26	1.30
red	14	1.21	0.59	0.60	0.24	5.44	3.21
yellow	15	1.08	0.61	0.60	0.33	2.10	0.90
magenta	15	1.39	0.60	0.69	0.28	0.71	0.63
cyan	15	2.42	1.30	1.58	0.79	4.25	3.42
white 9.5 (.05 D)	15	2.67	1.80	2.15	1.70	1.97	1.82
neutral 8 (.23 D)	15	1.66	0.80	1.95	1.15	2.43	2.91
neutral 6.5 (.44 D)	15	1.11	0.44	1.22	0.57	1.76	1.44
neutral 5 (.70 D)	16	1.40	0.71	1.75	0.97	1.28	1.73
neutral 3.5 (1.05 D)	15	1.33	0.56	1.66	0.84	1.47	1.98
black 2 (1.5 D)	15	1.93	1.11	2.38	1.50	3.32	3.12

Table 5.1: Color differences of the calibration scene.

Tables 5.1-5.3 provide detailed experimental results for the three test scenes. Each row corresponds to a particular reference color on the target. On the left hand side, the tables list five columns, providing the number of individual images that show the particular reference patch (not all images contain the full reference target), and mean μ and standard deviation σ for both color metrics. This evaluation provides quantitative results w.r.t. the homogeneity of our estimation of $P_\rho(\boldsymbol{p}\prime)$ in many different images of the same patch. The last two columns on the right hand side provide the difference of the average patch colors w.r.t. the known ground truth of the color reference target. One clearly sees an excellent homogeneity (i.e. small μ and σ on the left hand side), and increasing mean errors for increasing complexity of the scenes (i.e. ΔE on the right hand side). Table 5.2 further shows a case of slight overexposure, with the rows “white 9.5” and “neutral 8” indicating

	#img	$\mu(\Delta E^*_{ab})$	$\sigma(\Delta E^*_{ab})$	$\mu(\Delta E_{\text{CIE00}})$	$\sigma(\Delta E_{\text{CIE00}})$	$\Delta E^*_{ab,GT}$	$\Delta E_{\text{CIE00, GT}}$
dark skin	10	1.60	0.53	1.24	0.46	3.21	2.02
light skin	12	1.77	0.70	1.19	0.47	5.69	3.24
blue sky	12	1.79	0.83	1.60	0.83	0.54	0.53
foliage	12	1.62	0.67	1.02	0.42	3.25	2.29
blue flower	12	1.41	0.51	1.08	0.43	1.92	1.16
bluish green	12	1.77	0.91	0.94	0.43	1.84	0.85
orange	10	2.42	1.36	1.32	0.78	12.47	3.03
purplish blue	12	2.01	0.91	1.26	0.54	5.21	2.54
moderate red	11	1.76	1.15	0.88	0.56	1.60	0.89
purple	11	1.91	0.83	0.99	0.37	4.16	2.02
yellow green	11	1.87	0.92	0.75	0.35	6.64	2.32
orange yellow	11	1.39	0.72	0.77	0.37	7.29	2.13
blue	7	1.73	0.84	0.80	0.35	11.90	4.40
green	9	2.20	0.88	1.16	0.51	3.25	1.38
red	9	2.26	1.07	1.04	0.46	5.71	3.33
yellow	9	2.48	1.39	1.40	0.78	3.20	1.77
magenta	9	0.88	0.29	0.51	0.23	1.36	1.10
cyan	9	2.01	0.98	1.20	0.59	4.20	3.78
white 9.5 (.05 D)	6	48.33	36.30	16.10	11.82	64.40	31.92
neutral 8 (.23 D)	8	17.31	10.82	12.18	7.56	16.76	16.87
neutral 6.5 (.44 D)	8	1.71	0.77	2.07	1.04	0.61	0.79
neutral 5 (.70 D)	8	1.97	0.79	2.47	1.13	1.36	1.98
neutral 3.5 (1.05 D)	8	1.64	0.72	1.94	0.97	2.62	3.56
black 2 (1.5 D)	8	1.45	0.73	1.44	0.73	5.57	6.49

Table 5.2: Color differences of the daylight scene.

saturation.

A comparison of tables 5.3 and 5.4 illustrates the benefits of including step III in the radiometric processing scheme of section 5.2. Results including the simulation of incident light (table 5.3) are significantly better than without this processing (table 5.4).

Finally, table 5.5 summarizes the quantitative results of all four experiments by calculating mean values of all color patches, where we clearly see the improvement when excluding the two overexposed patches from the evaluation of the daylight scene.

5.3.2 Qualitative validation

Our main application domain of documenting rock panels with prehistoric petroglyphs is not very colorful. To demonstrate the general applicability of our method, I provide one qualitative example from a different domain. A colored, wooden statue of approx. 80cm height was scanned from 188 scanner positions. Figure 5.6a shows the dense 3D point cloud rendered without radiometric calibration, and fig. 5.6b a zoomed detail of the same object including radiometric calibration. These results are encouraging in various respects. First, the scanner exhibits sufficient depth-of-field so that even the deep folds of the cloth are in focus. Second, a qualitative visual comparison of the appearance of the real object

	#img	$\mu(\Delta E^*_{ab})$	$\sigma(\Delta E^*_{ab})$	$\mu(\Delta E_{\text{CIE00}})$	$\sigma(\Delta E_{\text{CIE00}})$	$\Delta E^*_{ab,GT}$	$\Delta E_{\text{CIE00, GT}}$
dark skin	23	1.94	1.02	1.28	0.63	8.77	4.54
light skin	24	2.89	1.52	2.19	1.27	6.09	3.03
blue sky	22	2.27	0.99	1.65	0.84	5.62	5.13
foliage	21	1.95	0.95	1.13	0.54	7.38	3.99
blue flower	22	2.45	1.47	1.83	1.31	5.72	3.78
bluish green	19	1.64	0.58	1.05	0.44	7.74	5.08
orange	29	5.71	3.79	3.50	2.35	18.21	5.74
purplish blue	26	2.54	1.46	1.11	0.49	11.09	7.24
moderate red	25	2.01	1.05	1.19	0.71	7.39	3.26
purple	23	1.82	0.98	1.07	0.70	4.56	2.47
yellow green	22	5.80	5.64	3.32	3.21	8.80	2.21
orange yellow	17	2.31	1.40	1.35	0.81	6.67	1.61
blue	31	2.04	0.87	1.05	0.46	9.18	4.12
green	30	3.14	1.49	1.52	0.72	10.56	3.98
red	26	2.68	1.63	1.20	0.68	10.51	5.33
yellow	26	2.13	1.43	1.16	0.77	1.53	0.51
magenta	23	2.57	1.48	1.39	0.99	8.36	4.63
cyan	17	1.82	0.90	1.31	0.72	11.62	5.58
white 9.5 (.05 D)	32	2.46	1.18	2.20	1.02	8.26	5.44
neutral 8 (.23 D)	31	2.81	1.67	2.24	1.14	1.52	1.34
neutral 6.5 (.44 D)	31	2.32	1.25	2.41	1.19	1.24	1.45
neutral 5 (.70 D)	28	2.22	1.37	2.43	1.47	3.45	4.03
neutral 3.5 (1.05 D)	22	1.61	0.89	1.72	0.99	4.13	4.98
black 2 (1.5 D)	17	1.92	0.86	1.99	1.05	6.85	7.34

Table 5.3: Color differences of the evaluation scene.

vs. the radiometrically calibrated reconstruction shows a clear benefit over the uncalibrated result.

5.3.3 Reconstruction of a colorful painting

Figure 5.7 shows to reconstruction of a reproduction by August Macke. ArcTron3D GmbH (`http://www.arctron.de/`) provided us the data. The reproduction was scanned with our scanner prototype. The final dense point cloud is made up of more than 39 million points.

5.4 Discussion

The presented approach allows an accurate reconstruction of radiometric surface property "in the wild" under Lambertian reflectance assumption. The resulting reconstruction consists of radiometric surface properties beyond photo texture in the sense that the incident illumination is cancelled out by the presented approach. Discussing about radiometric surface properties leads to a number of potential improvements:

	#img	$\mu(\Delta E^*_{ab})$	$\sigma(\Delta E^*_{ab})$	$\mu(\Delta E_{\text{CIE00}})$	$\sigma(\Delta E_{\text{CIE00}})$	$\Delta E^*_{ab,GT}$	$\Delta E_{\text{CIE00, GT}}$
dark skin	23	1.98	0.98	1.50	0.78	8.50	4.93
light skin	24	2.80	1.47	2.11	1.18	6.44	3.92
blue sky	22	3.26	1.78	2.88	1.73	6.89	6.69
foliage	21	2.81	1.31	2.16	1.17	8.23	5.35
blue flower	22	3.85	2.17	3.57	2.25	6.46	6.05
bluish green	19	4.82	2.73	3.58	2.18	6.85	4.89
orange	29	2.99	1.76	1.73	1.10	14.27	4.03
purplish blue	26	3.00	1.44	2.16	1.22	9.98	6.72
moderate red	25	3.25	1.87	2.44	1.55	8.10	5.90
purple	23	2.87	1.45	1.72	0.88	6.13	4.21
yellow green	23	5.26	3.26	3.56	2.31	8.79	6.71
orange yellow	18	4.63	2.96	3.00	1.93	6.14	3.96
blue	31	2.86	1.31	1.71	0.88	7.45	4.82
green	30	3.14	1.63	2.44	1.54	10.72	6.90
red	26	2.92	1.45	1.97	1.15	12.83	7.95
yellow	26	5.05	3.24	3.11	2.02	10.30	5.29
magenta	23	3.87	2.26	2.87	1.88	8.77	7.79
cyan	18	3.86	2.07	3.09	1.91	11.14	9.10
white 9.5 (.05 D)	32	3.89	2.28	2.92	1.44	9.12	6.50
neutral 8 (.23 D)	31	3.58	2.19	2.91	1.57	7.62	5.60
neutral 6.5 (.44 D)	31	3.39	1.89	3.30	1.67	6.28	5.31
neutral 5 (.70 D)	28	3.04	1.63	3.04	1.52	8.05	8.11
neutral 3.5 (1.05 D)	22	2.33	1.06	2.23	0.98	7.18	6.27
black 2 (1.5 D)	17	2.46	0.98	2.26	0.98	8.64	6.67

Table 5.4: Color differences of the evaluation scene without simulation of the actual incident light on the 3D surface points.

	$\mu(\mu(\Delta E^*_{ab}))$	$\mu(\sigma(\Delta E^*_{ab}))$	$\mu(\mu(\Delta E_{\text{CIE00}}))$	$\mu(\sigma(\Delta E_{\text{CIE00}}))$	$\mu(\Delta E^*_{ab,GT})$	$\mu(\Delta E_{\text{CIE00, GT}})$
Calibration scene	1.53	0.75	1.13	0.59	3.74	2.04
Daylight scene	4.39	2.73	2.31	1.34	7.28	4.18
Daylight scene without *white 9.5* and *neutral 8*	1.80	0.84	1.23	0.58	4.25	2.35
Evaluation scene	2.54	1.49	1.72	1.02	7.30	4.03
Evaluation scene without shading	3.41	1.88	2.59	1.49	8.54	5.99

Table 5.5: Summary of the color differences for the calibration scene, daylight scene, daylight scene without considering the patches *white 9.5* and *neutral 8*, evaluation scene, and the evaluation scene without simulation of the actual incident light on the 3D surface points.

(a) **(b)**

Figure 5.6: Radiometrically uncalibrated (a) and calibrated (b) 3D point clouds of a colored wooden statue (see text).

Extension to non-Lambertian reflectance: As already pointed out in chapter 2 one can find related work that is able to capture surface reflectance properties beyond Lambertian reflectance. This is a direction in which my work should be extended.
Despite this promising direction one has to overcome several challenges:

- The choice of an appropriate reflectance model: As mentioned at the beginning of this chapter, surface reflectance could be modelled with the *bidirectional reflectance distribution function* (BRDF). A BRDF encodes how light is reflected depending on the direction of the light source and the observer direction $\boldsymbol{\omega_o}$. One has to keep in mind that a BRDF only specifies the reflectance characteristic of "one" surface point. If the BRDF is varying across the surface, *e.g.*. the surface consists of different materials, one has to capture and model different BRDFs for different regions of the

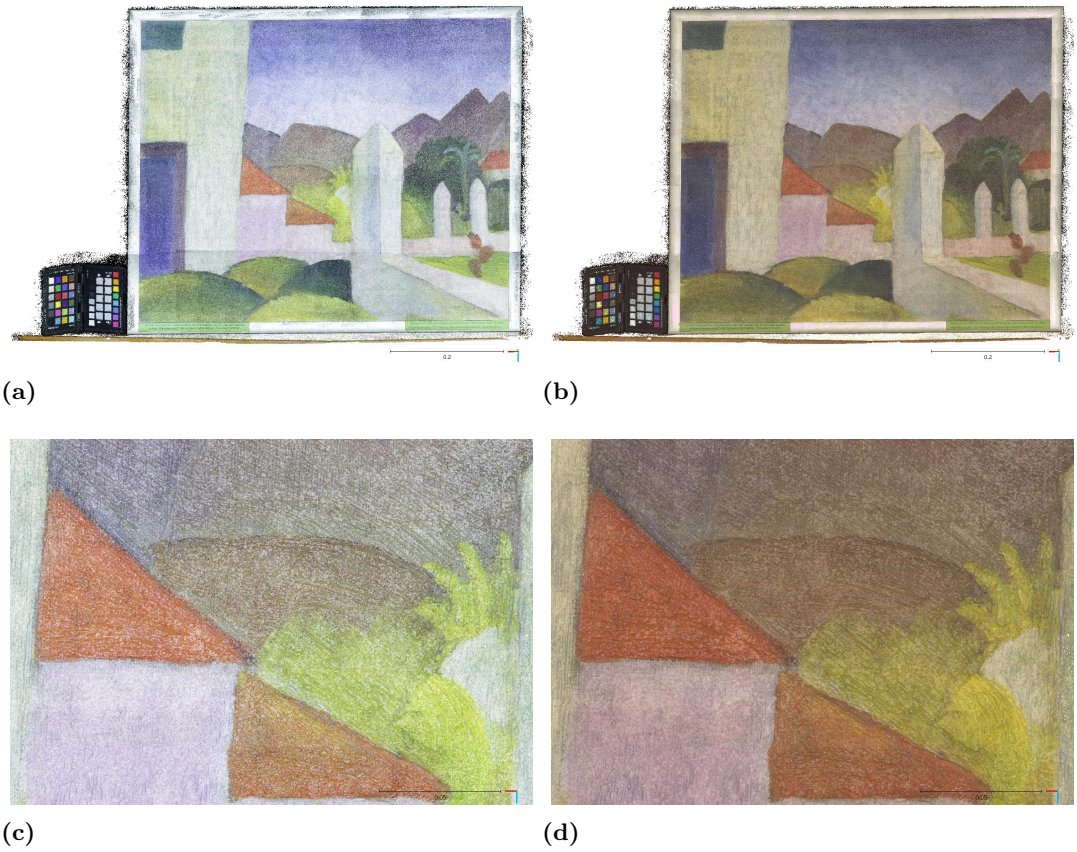


Figure 5.7: Reconstruction of a reproduction of a painting by August Macke. The images for this reconstruction were taken by ArcTron3D GmbH with our scanner prototype. (a) shows an initial dense point cloud before the estimation of the radiometric surface properties. In (b) one can see the result of the processing discussed in this chapter. (c) and (d) show a close-up of (a) and (b). One can see that the radiometrically accurate reconstruction exhibits much less noise than the initial dense point cloud.

surface - this sort of BRDF is called *spatial-varying BRDF* (SV-BRDF). The main challenge is a proper sampling of the spatial and angular domain of the SV-BRDF which is a six-dimensional function (two dimensions encode the spatial position on the surface and the four additional dimensions encode the angular dependencies).
To simplify this challenging sampling task one could use parametric BRDF models like, [29], [16], etc. and fit a sparse set of sampled angular data. Another direction would be to restrict the angular sampling to a certain illumination-view-point configuration and use a data driven BRDF model to infer the missing and unknown data. This approach was *e.g.* taken by [56]. Another way would be to find a subset of pairs of illumination directions and observer directions that best characterize the

surface reflectance and sample just this subset. In a next step one has to infer the missing dimensions based on this subset. Filip *et al.* [44] have done work in this direction where they use a robotic manipulator to position the light source and the camera in an appropriate configuration (which was done in a lab).
These potential extensions approximate the true reflectance behaviour of the surface. For the setting of the 3D Pitoti project, and in my opinion this is the case for almost all practical use cases, the direct sampling of an SV-BRDF is prohibitive because of the time required to sample this function.

- Fast acquisition of necessary reflectance properties "in the wild": In my opinion, this is the main challenge when one wants to capture radiometric surface properties beyond Lambertian reflectance. The *Parthenon-project* [9] reflects very well the huge amount of time-consuming work. The BRDF measurement [149] for example was done during the night and just for a representative part of the *Parthenon*-temple.
  In the work of Georgoulis *et al.* [56] the authors propose a method where shape and reflectance are estimated jointly. To infer the reflectance the authors take an image with a characteristic incident illumination and observer direction configuration. The obtained reflectance data are further extrapolated using the *MERL BRDF measurements* [96] as prior information. In this way the authors could acquire the necessary data in a fast way. As far as this paper suggests the data acquisition was done inside a laboratory and not "in the wild".
- We tried to use the existing possibility of our scanner prototype to measure different incident light directions $\boldsymbol{\omega_i}$. Triggering each of the 22 LED groups of our illumination provides us with 44 different directions per scanner position (22 per camera). But the resulting investigation [146] showed that the benefits of this concept are rather limited. First, we have to go back to a dark laboratory, because 10 LEDs are not sufficient bright to benefit from our flash-no-flash idea. Second, the individual groups are spatially quite spread (see the LED arrangement on the circuit board in figure 4.6), so that we do not obtain a particular crisp direction $\boldsymbol{\omega_i}$. Third, the given focal length, working distance, and size of the illumination limits the variability of angular sampling.

Using a GPU to speed up the computation time: Simulating the incident illumination for a single 16 million pixel image is a time consuming task. In the current setting I use multiple cores to process this computation. Using a GPU would accelerate this computation even more because of the higher number of cores and the specialized architecture for computing light-surface interaction.

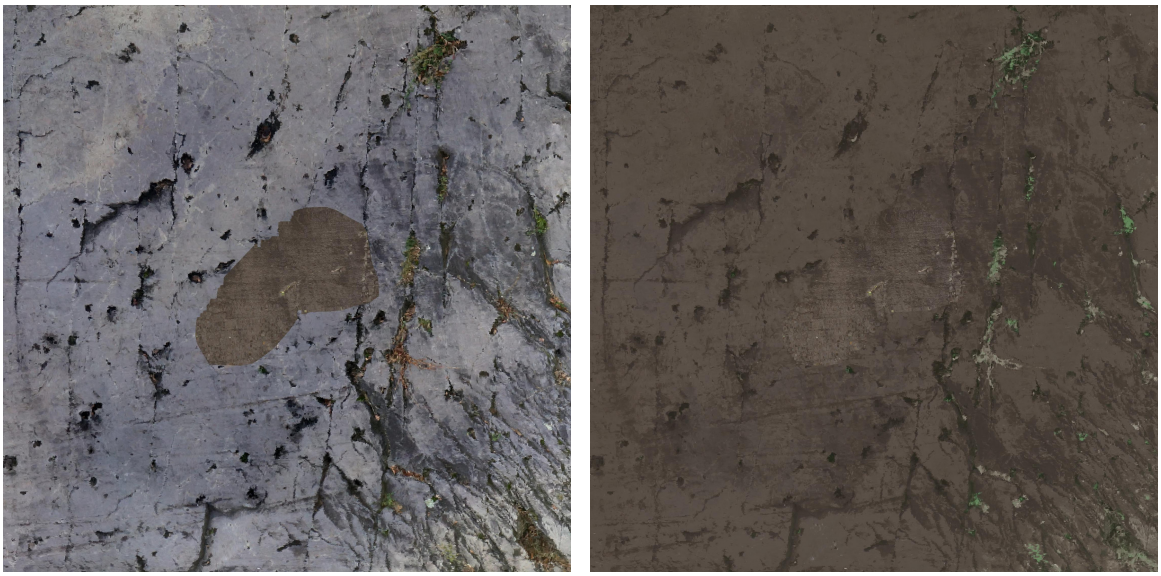
Using a render software to compute a more accurate surface illumination: One drawback of the presented approach to simulate the incident light is the fact that no self shadowing is handled. A ray-tracing software could handle this effect. Open source software like PBRT [117] is available which can also handle complex reflectance properties.

6

Radiometry propagation

In chapter 5 I discussed my approach to estimate accurate radiometric surface properties. In the case of the 3D Pitoti project we were confronted with large scale scenes (compared to the scale at which the scanner operates). For such sites, "true" radiometry for the complete 3D point cloud is very expensive to obtain. This chapter presents a method that is capable to estimate the radiometric surface properties of an entire scene despite the fact that we only have access to the "true" radiometry of a small part of it. Figure 6.1 illustrates the results and motivation of the presented approach.

Figure 6.2 outlines the proposed method in more detail. First, it is assumed to have a possibly large-scale 3D reconstruction at a certain spatial resolution. Regarding its surface properties I assume that the images were taken under diffuse illumination conditions resulting in a reconstruction colored with the photo texture of the images, but without



(a) Init **(b)** Final

Figure 6.1: 3D point clouds, one obtained by an MAV based reconstruction (large, grey area) and the second is obtained by the prototype scanner. (a) shows part of Seradina 12C, reconstructed at low spatial resolution using an MAV, plus a smaller area (Aerea345) that has been scanned at high spatial resolution and true radiometry. I first transfer the true radiometry to the spatially registered subregion of the low resolution point cloud, and then propagate this radiometry to the remainder of the reconstruction, as shown in (b).

direct, cast shadows. Recent advance in the field of autonomous image capture for photogrammetry, e.g. Mostegel *et al.*[104], allows the reconstruction of large-scale scenes with minimal effort. However, the color values of such a reconstruction are still calculated based on the captured images and hence, no actual radiometric surface color is available. I used the MAV based reconstruction (see section 3.3.1) obtained by project partners for this point cloud and the grey area in figure 6.1a shows a section of this reconstruction.
Second, I assume that I have access to an accurate radiometric reconstruction of parts of the same scene, possibly captured at a different, more detailed spatial scale. Here, the concept builds on the idea, that several *radiometrically relevant* parts of the scene have been scanned - it would be unrealistic to try to solve the problem for surfaces exhibiting completely different radiometric properties than the samples taken. To obtain such point clouds I used the scanner prototype. In figure 6.1a I show the radiometrically accurate reconstruction of *AREA345* (see section 3.3.3) obtained during the 3D Pitoti project.
The accurate radiometric values are mapped to the entire 3D scene in a two-stage process: First, I transfer the "true" radiometric values to the photo texture of the spatially corresponding part of the scene, and second, I propagate these values to the entire scene

using affinity information.

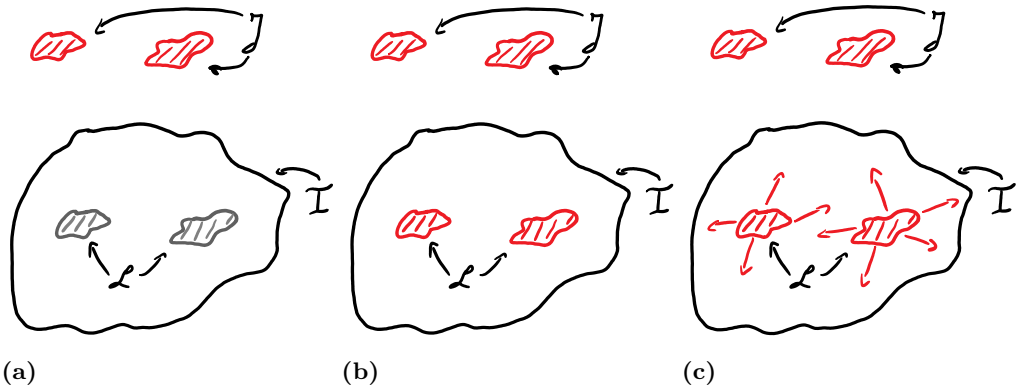


Figure 6.2: Description of my method. (a) sketches the initial setting. Two types of point clouds are available: $\mathcal{I}$, which is colored with photo texture and $\mathcal{J}$, for which radiometric surface properties are available. The set of points in $\mathcal{L}$ is the set of matches between the spatial coordinates of $\mathcal{I}$ and $\mathcal{J}$ ($\mathcal{J}$ is drawn above $\mathcal{I}$ for visualization purpose, actually they are registered). (b) shows the result after color transfer trough Optimal Transport as described in section 6.1. I transfer the color distribution from point cloud $\mathcal{J}$ to the points in the set $\mathcal{L} \subset \mathcal{I}$. (c) sketches the color propagation approach in which I propagate the known radiometric values to the areas where only photo texture is available. This process is described in section 6.2.

6.1 Color Transfer through Optimal Transport

Color transfer was originally used to transfer the color characteristic of one image to another one. Seminal work about color transfer between two images was published by Reinhard *et al.*[127], in which the authors use a color space that minimizes the correlation between color channels [134] and simple statistical analysis to transfer the color characteristic between two images. Pitié *et al.*([119] and [120]) transform one N-dimensional color distribution into another N-dimensional distribution by iterativly calculating their marginal distribution with subsequent histogram specification [60]. This provides more flexibility than [127] because one is not restricted to a certain distribution. In [126], Rabin *et al.* established a link between this prior work on distribution transfer and the *Wasserstein metric*.

This section explains how color values are transferred between spatially overlapping areas were both is available: the "true" radiometric surface properties and the photo texture. This process must be able to handle inaccurate spatial registration as well as potentially different spatial scales of the two 3D point clouds. Therefore, I prefer not to depend on

the spatial correspondence but instead transfer the global color characteristic between the patch with "true" radiometry and the corresponding patch of the photo-textured data.

The set $\mathcal{I}$ is defined as the point cloud where only the photo texture is known, the set $\mathcal{J}$ as the point cloud where radiometric surface properties are available, and the set $\mathcal{L} = \texttt{match}(\mathcal{I}, \mathcal{J})$ as the set of matches between the spatial coordinates of point cloud $\mathcal{I}$ and $\mathcal{J}$. This means that the set $\mathcal{L}$ contains the points of $\mathcal{I}$ ($\mathcal{L} \subset \mathcal{I}$) for which the radiometric surface properties are known but because of different scales and inaccurate registration, no point-wise match between $\mathcal{J}$ and $\mathcal{L}$ is possible. Further, $\mathbf{r}_j = [r_r, r_g, r_b]$ is denoted as the "true" radiometry of a point $j \in \mathcal{J}$ expressed in RGB-color space and $\mathbf{p}_i = [p_r, p_g, p_b]$ as the corresponding RGB-vector of the photo texture of a point $i \in \mathcal{I}$. I represent the two matrices $\mathbf{R} = \{\mathbf{r}_j\}_{j \in \mathcal{J}}$ and $\mathbf{P} = \{\mathbf{p}_i\}_{i \in \mathcal{L} \subset \mathcal{I}}$ as two distributions in three dimensional color space and want to transform the distribution $\{\mathbf{p}_i\}$ such that it is as similar as possible to the distribution $\{\mathbf{r}_j\}$.

I utilize the concept of *Optimal Transport* (OT) [155] to achieve this desired color transformation. To simplify the explanation of this concept let us assume that the number of points in $\{\mathbf{r}_j\}_{j \in \mathcal{J}}$ and $\{\mathbf{p}_i\}_{i \in \mathcal{L} \subset \mathcal{I}}$ is equal, where L denotes the number of points. In OT one minimizes the cost of transporting one distribution onto another distribution using an assignment σ:

$$\min_{\sigma \in \Sigma_L} \sum_{i \in \mathcal{L}} d_{Lab} \left(\mathbf{p}_i, \mathbf{r}_{\sigma(i)} \right)^2 \ , \tag{6.1}$$

where Σ_L is the set of all permutations of L elements and d_{Lab} is the distance between two color values in the *Lab*-color space. The solution of this optimization problem is an assignment which maps the photo texture $\mathbf{t}_i$ of point i to a radiometric surface color $\mathbf{r}_{\sigma(i)}$, expressed as

$$\mathbf{p}_i \mapsto \mathbf{r}_{\sigma(i)} = \hat{\mathbf{r}}_i := [\hat{r}_r, \hat{r}_g, \hat{r}_b] \ , \tag{6.2}$$

where $\hat{\mathbf{r}}_i$ is the result of this mapping. One could use the flow obtained by calculating the *Earth Movers's Distance* [133] to find such a mapping but it suffers from the high ($O(n^3 \log n)$) computational complexity [142]. Therefore, the mapping is computed using the *Sliced Wasserstein Distance* [126] approximation in which the *Wasserstein metric* is approximated by a set of random 1D projections of the distributions and calculating 1D assignments.

Rabin *et al.*define the sliced Wasserstein distance as follows

$$\widetilde{W}(\mathbf{P}, \mathbf{R}) = \int_{\mathbf{\Theta}} \underbrace{\min_{\sigma_\Theta} \sum_i \left(\langle \mathbf{p}_i - \mathbf{r}_{\sigma_\Theta(i)}, \mathbf{\Theta} \rangle \right)^2}_{\mathcal{E}(\mathbf{P})} d\mathbf{\Theta} \ , \tag{6.3}$$

in which $\mathbf{\Theta}$ is a 1D line onto which $\mathbf{p}_i$ and $\mathbf{r}_{\sigma_\Theta(i)}$ are projected. In equation 6.1 the distance

between colors is calculated in the *Lab*-color space. For consistency, for the remainder of this section $\{\mathbf{p}_i\}$ and $\{\mathbf{r}_i\}$ are converted into this color space. The goal is to transform $\mathbf{P}$, making it as similar as possible to $\mathbf{R}$, to minimize $\widetilde{W}(\mathbf{P}, \mathbf{R})$, which approaches a minimum, if, for every $\mathbf{\Theta}$, $\mathcal{E}(\mathbf{P})$ is minimized. Because of the 1D projection in $\mathcal{E}(\mathbf{P})$ one can minimize this term w.r.t. $\mathbf{P}$ by calculating a *Histogram Specification* [60] of the projected points

$$\{r_{\Theta,i}\} = \texttt{hist\_spec}(\text{proj}_{\mathbf{\Theta}}\mathbf{P}, \text{proj}_{\mathbf{\Theta}}\mathbf{R}) \ , \tag{6.4}$$

where the operator $\text{proj}_{\mathbf{\Theta}}$ projects points onto the 1D line $\mathbf{\Theta}$, and $\texttt{hist\_spec}(\{x_i\}, \{y_i\})$ performs histogram specification such that $\{y_i\}$ matches $\{x_i\}$ as closely as possible (which also holds for different numbers of points in $\{x_i\}$ and $\{y_i\}$). Minimizing 6.3 with respect to $\mathbf{P}$

$$\mathbf{P}^\star = \{\hat{\mathbf{r}}_i\} = \underset{\mathbf{P}}{\text{argmin}}\, \widetilde{W}(\mathbf{P}, \mathbf{R}) \ , \tag{6.5}$$

can be done using *Stochastic Gradient Descent* [20] where the 1D line $\mathbf{\Theta}$ is randomly generated. The update rule for $\mathbf{P}$ reads

$$\mathbf{P}^{t+1} = \mathbf{P}^t + \gamma \nabla_{\mathbf{P}} \mathcal{E}(\mathbf{P}) \ , \tag{6.6}$$

where γ is the step size and $\nabla_{\mathbf{P}}\mathcal{E}(\mathbf{P})$ is the gradient of $\mathcal{E}(\mathbf{P})$. Computing this gradient

$$\nabla_{\mathbf{p}_i}\mathcal{E}(\mathbf{P}) = 2\mathbf{\Theta}\left(\mathbf{\Theta}^T\mathbf{p}_i - \underbrace{\mathbf{\Theta}^T\mathbf{r}_{\sigma_\theta(i)}}_{r_{\Theta,i}}\right) \ , \tag{6.7}$$

involves solving the histogram specification eq. 6.4 and can be done very efficiently. The stochastic gradient descent algorithm terminates after a predefined, fixed number of iterations, or as soon as no further changes occur.

6.2 Radiometry propagation

Why do I need to treat radiometry propagation different from color transfer discussed in section 6.1? Because I cannot rely on patches where I have both, "true" radiometry $\mathbf{r}$ and phototexture $\mathbf{p}$, but wish to propagate true radiometry into the unknown. In this case, one may observe similar phototexture, where it will make sense to propagate according radiometric values, but one also may observe significantly different phototexture, which has not been covered by the "true radiometric" scans at all. These values should not be included into the color distributions used in section 6.1. I therefore require a different method to solve this case and explain my solution, which is based on label propagation,

below.

The result of color transfer described in the previous section is a set $\{\hat{\mathbf{r}}_i\}_{i\in\mathcal{L}}$ of radiometric values expressed as RGB color vectors. This set of vectors is associated with a subset of color values of the initial photo textured point cloud $\mathcal{I}$. The goal is to propagate the available radiometric values to those areas $\{\mathbf{p}_i\}_{i\in\mathcal{I}\setminus\mathcal{L}}$ in the point cloud where I have no information about the radiometry. I use ideas from the field of semi-supervised learning, namely *Label Propagation* (LP) by Zhu and Ghahramani [179], to propagate the radiometric values to the entire scene. For this sake I first construct a matrix $\hat{\mathbf{R}}_l = \{\hat{\mathbf{r}}_i\}_{i\in\mathcal{L}}$ containing the known (labeled) radiometric values and second, a matrix $\hat{\mathbf{R}}_u = \{\mathbf{p}_i\}_{i\in\mathcal{I}\setminus\mathcal{L}}$ containing the photo texture (unlabeled). Next, I concatenate both matrices to obtain $\hat{\mathbf{R}} = (\hat{\mathbf{R}}_l, \hat{\mathbf{R}}_u)$. In the same way color values in the point cloud $\mathcal{I}$ are ordered according to the criterion, whether a radiometric color value can be associated or not.

To set up the LP method, an affinity matrix $\mathbf{W}$ consisting of Gaussian kernels is constructed:

$$w_{ij} = \exp\left(-\frac{d_{Lab}(\mathbf{p}_i, \mathbf{p}_j)^2}{2\sigma_{Lab}^2}\right) \ , \tag{6.8}$$

in which i and j are points from the point cloud $\mathcal{I}$, d_{Lab} is the distance in the *Lab*-color space, and σ_{Lab} is the width of the Gaussian kernel. To illustrate the concept behind $\mathbf{W}$, it can be interpreted as a graph whose nodes are the points in $\mathcal{I}$, and the edge weights are proportional to the color similarity between points in $\mathcal{I}$. Because of the high storage requirements and computational burden of such an affinity matrix the computation of the affinity is limited to the k nearest neighbours in spatial and radiometric domain, by considering only a k-neighbourhood for every point i

$$j \in \mathsf{N}_{k,Lab}(\mathbf{p}_i) \cup \mathsf{N}_{k,spatial}(\mathbf{p}_i) \ , \tag{6.9}$$

where $\mathsf{N}_{k,Lab}$ denotes the set of radiometric neighbours in the *Lab*-color space, $\mathsf{N}_{k,spatial}$ the set of spatial neighbours, and k is the number of neighbours to consider. The diagonal degree matrix $\mathbf{D}$ is constructed according to

$$d_{ii} = \sum_{j\in I} w_{ij} \ . \tag{6.10}$$

Using $\mathbf{W}$ and $\mathbf{D}$, the probabilistic transition matrix $\mathbf{T}$ is defined as

$$\mathbf{T} = \mathbf{D}^{-1}\mathbf{W} \ . \tag{6.11}$$

The element t_{ij} of $\mathbf{T}$ corresponds to the probability of jumping from node j to node i. According to Zhu and Ghahramani [179] the radiometric values (labels) can be propagated

σ_{Lab}	1.0
neighborhood size of $\mathsf{N}_{k,Lab}$	15
neighborhood size of $\mathsf{N}_{k,spatial}$	15
number of iterations	2500

Table 6.1: Parameters for the experiments of section 6.3.1.

through the graph by the following algorithm:

- Initialize: $\hat{\mathbf{R}}^t = (\hat{\mathbf{R}}_l^0, \hat{\mathbf{R}}_u)$ with $\hat{\mathbf{R}}_l^0 = \hat{\mathbf{R}}_l$
- Iterate:
  1. Propagate: $\hat{\mathbf{R}}^{t+1} = \mathbf{T}\hat{\mathbf{R}}^t$
  2. Persist in known radiometry: $\hat{\mathbf{R}}_l^{t+1} = \hat{\mathbf{R}}_l^0$

Step 1 propagates the radiometric values to their neighbourhood, and step 2 ensures that the initial "true" radiometry does not fade out. This algorithm terminates either after a predefined, fixed number of iterations, or as soon as no further changes occur (i.e. $\hat{\mathbf{R}}_u^{t+1} = \hat{\mathbf{R}}_u^t$). Finally, the new, propagated color values in $\hat{\mathbf{R}}_u^{t+1}$ are stored in the corresponding points in $\mathcal{I}$.

6.3 Experiments

This section deals with the validation of the proposed approach on three different datasets to show its general applicability.

6.3.1 Radiometric correction of a large-scale cultural heritage reconstruction

For this experiment I consider the real world archaeological use case within the 3D Pitoti project [1].

I have two types of reconstructions at hand: first, the large-scale reconstruction (`Ser12c`, see section 3.3.1) based on images obtained by a micro aerial vehicle and second, a set of radiometrically corrected small-scale, but high detail, reconstructions (`A`, `B`, `C`, `D` = AREA345, see section 3.3.3) (see figure 6.3). The reconstructions `A`, `B`, `C`, and `D` were scanned by the scanner prototype and calculated by the presented approach in chapters 4 and 5. In figure 6.4, I show a close-up of a section of reconstruction `D` to illustrate its high spatial resolution.

Given these reconstructions I perform first the color transfer as explained in section 6.1 and second, the radiometry propagation described in section 6.2.

[1] I thank the authors of [104] for providing their data.

(a) (b) (c)

Figure 6.3: Images of the used dataset for the radiometric correction of 3D point clouds. (a) shows the large-scale photo textured `Ser12c` reconstruction. (b) shows the small-scale, high resolution, reconstructions with "true" radiometry (top row: `A`, `B`, `C`, bottom row `D`). (c) visualizes in red the registration of `A`, `B`, `C` and `D` w.r.t. `Ser12c`. The number of 3D points for the individual reconstructions is as follows: `Ser12c`: 21 million, `A`: 63 million, `B`: 23 million, `C`: 52 million, and `D`: 273 million points. Figure 6.4 shows a close-up of a section of the reconstruction `D`.

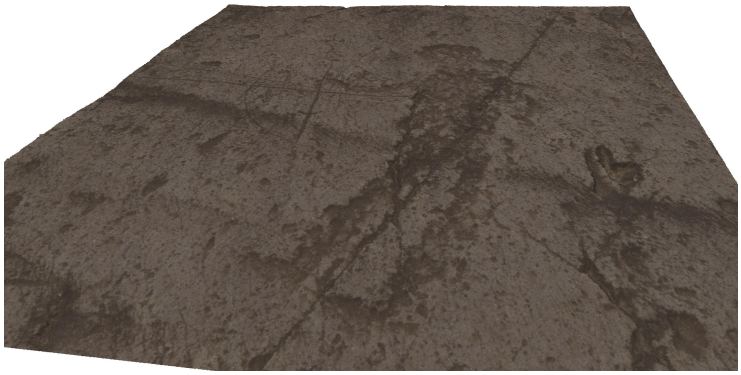
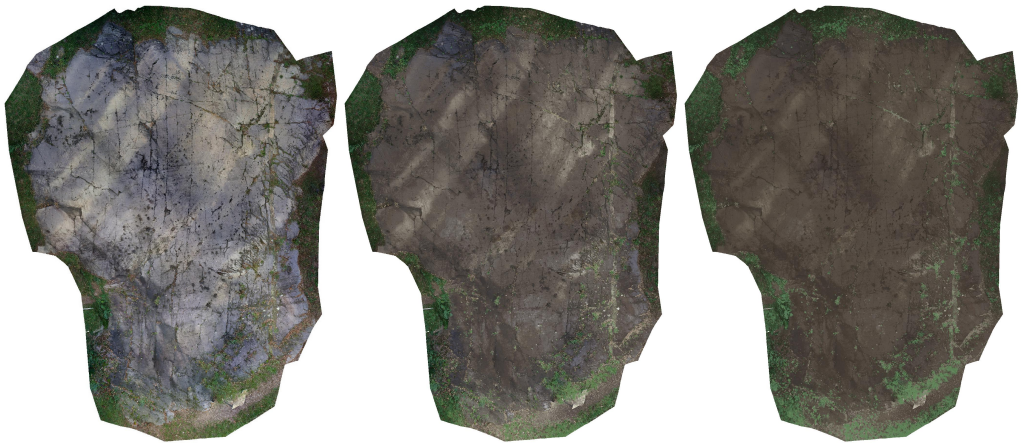


Figure 6.4: 3D reconstruction of an individual petroglyph showing a hunter with bow. The point cloud consists of 21.7 million points and the petroglyph covers a tiny area of approx. 10cm^2 on the rock.

omitted data:	`A`	`B`	`C`	`D`
d_{Lab}:	5.95	3.58	5.96	5.70

Table 6.2: Leave-one-out evaluation of radiometry propagation on the 3D point cloud.

Figure 6.5 shows the initial point cloud `Ser12c` before (a) and after (c) radiometric correction (see parameter settings for this experiment in table 6.1). In addition to this qualitative, visually pleasing result, I provide a quantitative leave-one-out evaluation shown in



(a) (b) (c)

Figure 6.5: This figure shows the progress of the radiometric value propagation explained in section 6.2. (a) shows the initial 3D point cloud. (b) shows an intermediate result (after 700 iterations) and (c) the final result at iteration 2500.

table 6.2: I correct `Ser12c` by using just three out of four "true" radiometry reconstructions and use the remaining one to validate the performance. As performance measure I use the mean Euclidean distance in *Lab*-color space.
The computation of 2500 iterations requires 4.5 hours (21 million points, $\mathsf{N}_{k,Lab} = 15$, and $\mathsf{N}_{k,spatial} = 15$, single core).

6.3.2 Radiometric correction of a colorful painting

One might argue that the point clouds processed in section 6.3.1 are not very colorful. Therefore, this section provides an example for the 3D point cloud of a colorful painting. Figure 6.6 shows the experimental setting. I use the same reproduction of a painting by *August Macke* as in section 5.3.3 see figure 5.7 The data include: A photo-textured 3D point cloud of the whole painting, three point clouds representing subregions of the painting with radiometrically accurate color values (see fig. 6.6.a for the phototextured pointcloud with the three subregions superimposed), and a complete, radiometrically accurate reconstruction of the painting that serves as ground truth to validate our approach (see fig. 6.6.c).
As described in section 6.1, I first transfer the radiometric values onto the photo-textured point cloud. Next, the radiometric values are propagated across the photo-textured reconstruction (as explained in section 6.2). Figure 6.6.b shows the result. Parameter settings for this experiments are as follows: $\mathsf{N}_{k,Lab} = 15$, $\mathsf{N}_{k,spatial} = 15$, and 2500 iterations. In comparison with the ground truth point cloud, I obtain a mean Euclidean distance in

(a) (b) (c)

Figure 6.6: Radiometric correction of a colorful painting. (a) shows the initial point cloud with three regions of "true" radiometry superimposed. (b) shows the final result after radiometry propagation and (c) the ground truth. The mean Euclidean distance in *Lab*-color space between the final result and the ground truth is $d_{Lab} = 6.92$.

Lab-color space of $d_{Lab} = 6.92$.

This experiment clearly demonstrates the benefits of the method. If small portions of all radiometrically meaningful colors of a scene are scanned, these "true" colors can successfully be propagated to a potentially much larger phototextured 3D point cloud.

6.3.3 Quantitative validation of radiometry propagation in the image domain

In this section, I provide further quantitative results for our radiometry propagation approach (section 6.2) on the image dataset provided by Gehler *et al.*[55] and reprocessed by Lynch *et al.*[92]. This dataset consists of 482 3.2MPixel images of outdoor and indoor scenes including a color reference target. Based on this reference target, I apply color correction [164] and define the corrected images as the set of images with "true" radiometry. For each of these images, I randomly sample a fraction of its pixel colors, except pixels within the color reference target, and use them as the initial, "true" radiometric values (i.e., known labels, see section 6.2). Table 6.4, shows three different types of quantitative performance indicators for several fractions of sampled pixel values: first, *Structural Similarity* $SSIM$ [157] which is tailored to match the characteristics of the human visual system; second, $PSNR$ in dB; and third, Euclidean distance in *Lab*-color space d_{Lab}. Pixels inside the color reference target are omitted from this validation. Figure 6.7 shows radiometry propagation results for a sample image, and the parameters used in these experiments are provided in table 6.3.

I am fully aware that random sampling of individual points of the "true" radiometric values is not a realistic setting. However, I conducted this experiment to obtain statistically meaningful quantitative results on a large number of images, varying the sample size

(a) Original image

(b) "True" radiometry

(c) Result for a sample size of 30%.

(d) Result for a sample size of 5%.

Figure 6.7: Radiometry propagation example on an image from [55]. For visualization purpose, the original (much darker) image (a) has been *Gamma* corrected. (b) shows the corresponding image with "true" radiometry, from which we randomly sample pixel values. Quantitative results for (c) are $PSNR = 35.45dB$, $SSIM = 0.92$, and $d_{Lab} = 3.14$. For (d), $PSNR = 30.89dB$, $SSIM = 0.852$, and $d_{Lab} = 5.56$. The pixel values of the color reference target are omitted for both, sampling and validation.

between 1% and 30% of the pixels. To conclude this experiment, I provide two examples for the more realistic scenario, where a supervisor decides which parts of an image need to be sampled to reconstruct the remaining part via radiometry propagation. I show experiments on two images of the dataset [55], `IMG_0284` and `IMG_0881`, where I manually selected two small areas of radiometrically "true" values and propagated them across the entire image. Figure 6.8 shows the selected images, sampled areas and the result of this user guided approach. Table 6.5 provides the validation results in terms of $PSNR$, $SSIM$, and d_{Lab} and the parameter settings used for radiometry propagation.

In summary, the experiments presented in section 6.3.3 provide quantitative validation of the radiometry propagation approach (sec. 6.2). The results show that "true" radiometry can be successfully propagated even if the sample size covers only a few percent of the data, where the method still works remarkably well for a sample size of 5%, but breaks

σ_{Lab}	1.0
neighborhood size of $\mathsf{N}_{k,Lab}$	25
neighborhood size of $\mathsf{N}_{k,spatial}$	4
number of iterations	120

Table 6.3: Parameter settings for the random sampling experiment in section 6.3.3.

	$PSNR$		$SSIM$		d_{Lab}	
sample size	mean	std	mean	std	mean	std
30%	35.05	3.55	0.92	0.05	3.41	1.41
20%	34.10	3.40	0.90	0.06	3.93	1.59
10%	32.42	3.27	0.88	0.07	4.80	1.84
5%	30.16	3.14	0.84	0.08	5.91	2.08
1%	15.68	3.62	0.51	0.12	20.44	6.58

Table 6.4: Results of radiometry propagation on the dataset provided by [55]. Sample size is given in % of pixels randomly sampled from the "true" radiometry image. Mean and standard deviation are calculated for the complete dataset of 482 images.

image name	$PSNR$	$SSIM$	d_{Lab}	$\mathsf{N}_{k,Lab}$	$\mathsf{N}_{k,spatial}$	num. of iterations
`IMG_0284`	26.88	0.81	5.93	125	4	1000
`IMG_0881`	27.14	0.77	8.58	125	4	2500

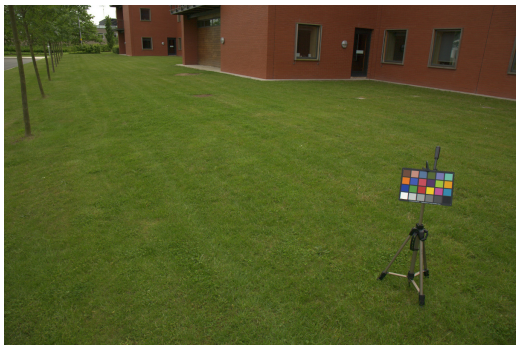
Table 6.5: Quantitative results of user guided radiometry propagation.

down for a sample size of 1%.

6.4 Discussion

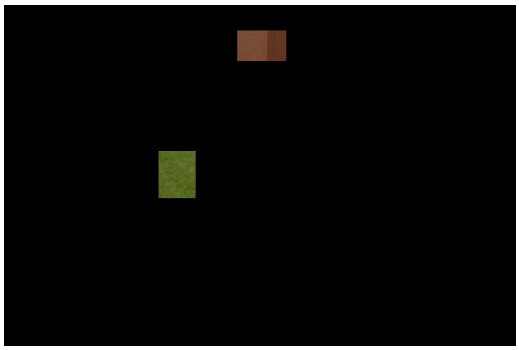
Chapter (chapter 5) presented my approach to estimate the radiometric surface property. Despite the fact that we can collect radiometric data in a fast way with the scanner prototype, reconstructing large scale radiometric surface properties is a very expensive and time consuming task. On the other hand, when one wants to reconstruct rock-art in its context, not every part of the scene is equally important. For example, in the 3D Pitoti project we observed that not the entire rock panel is covered with rock-art. Scanning the whole rock to reconstruct the radiometric surface properties of every single surface point would be highly inefficient and unnecessary. For such a setting, the approach of this chapter could be used.

The presented approach in this chapter allows to propagate known radiometric values into areas where only photo texture is available. I made some assumptions which could possibly be relaxed for further applicability of the presented approach:



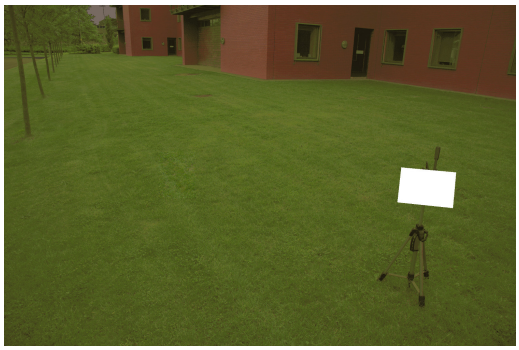
(a) "True" radiometry of `IMG_0284`

(b) "True" radiometry of `IMG_0881`



(c) Sampled areas

(d) Sampled areas



(e) Result $d_{Lab} = 5.93$

(f) Result $d_{Lab} = 8.58$

Figure 6.8: Two images (`IMG_0284` and `IMG_0881`) from the Gehler *et al.*[55] dataset used for radiometry propagation. (a) and (b) show the "true" radiometry from which I sample values according to the areas shown in (c) and (d). (e) and (f) show the result of radiometry propagation and the distance in *Lab*-color space (further quantitative measures can be found in table 6.5). As expected, regions with colors that are not covered by the manually selected areas are not well reconstructed, *e.g.*the white T-shirt or the magazines on the book shelf.

1. Diffuse illumination of the photo textured area. In my approach this is necessary to ensure that the pairwise color distance d_{Lab} (see equation 6.8) between two surface points is not distorted by varying incident illumination. A cast shadow might be such a distortion. One way would be to detect and correct such shadows in advance using approaches like the one of Shen *et al.* [140]. After shadow detection and correction one could use my approach as presented.
2. A user has to select and scan representative surface areas from which the propagation starts. A computational approach could guide the user to representative regions to scan.
3. The propagated information is the Lambertian surface reflectance. Optimal transport is also capable of interpolating between different BRDFs as work of Solomon *et al.* [144] suggests. Using this property one approach would be to capture the representative BRDF of a scene / object, transfer the BRDF using optimal transport, and propagate the BRDF parameters using label propagation.

During the experiments I observed that the *optimal transport* approach induced noise in the transported color distribution. This was already noticed in the optimal transport community and can be tackled using regularization. Ferradans *et al.* [41] proposed a regularized optimal transportation scheme for color transport specialized towards color transfer.

7

Conclusion

This thesis presented my work on accurate surface geometry and radiometry acquisition "in the wild". The context in which I did this work was the 3D Pitoti project, a European project on cultural heritage.
Besides cultural heritage, the presented approach and scanner prototype could also be used in other fields like forensics. Crime scenes could be scanned with high spatial resolution and accurate colors which would be beneficial.

In terms of the geometry I used a stereo-rig and a *strucutre from motion* approach in combination with a tachymeter to reconstruct the desired rock-art and its location in the landscape. The resulting high-resolution reconstruction is accurate, which I validate by comparing it with available ground truth data.

The presented constrained bundle adjustment approach is flexible to use. In my thesis, I used it to incorporate the stereo and tachymeter measurement constraints. But the approach is not limited to such a setting. Eg. if one has multiple cameras rigidly attached and knows the relative orientations between them this could be modelled as a constraint during bundle adjustment. Furthermore, the data from a GPS receiver could be incorporated because in its essence this is very similar to a tachymeter measurement.
However, incorporating different sensor modalities will result in an estimation process in which one has to specify the confidence in the sensor. In my work, I used the parameters ϵ_b (in equation 4.2) and ϵ (in equation 4.3) to specify this confidence. A more accurate model would be to model this with a probabilistic noise model, which is done in *sensor fusion* (*e.g.* [91]).

In terms of radiometry, I developed an approach that relies on the geometric reconstruction and on a calibrated custom built LED light. Interfering environmental illumination is cancelled out by taking images with and without illuminating the scene with the LED light and by subsequent frame differencing. The knowledge of the calibrated LED light is used to infer the Lambertian surface reflectance properties of the rock-art. I validate the accuracy of the radiometric reconstruction using a color reference target.
A definitively important and exciting topic for cultural heritage is the measurement or estimation of more complex surface reflectance properties. The scanner prototype is only of limited use for this task. The scanner can illuminate the scene from different directions but the range of directions is rather narrow. Furthermore, the LEDs are grouped in such a way that they exhibit a rather large area that will illuminate the scene. But when one wants to measure direction dependent reflectance characteristics, the light source should be very compact.
In my opinion the LED-only reflectance scanner presented by Ben-Ezra *et al.* [18] could be, with some modifications, used outdoors. On the other hand approaches like the multi-view shape and reflectance estimation from Oxholm and Nishino [116] could be incorporated in a standard photogrammetric reconstruction framework.

To tackle the problem of large scale reflectance acquisition I propose a method to propagate known reflectance into areas where only photo texture is available. The presented approach assumes that a user selects and scans areas of a scene with scene-characteristic reflectance behaviour, and that a photo-textured reconstruction of this scene is available. My approach transfers the scanned radiometry onto the photo-textured scene and subsequently propagates the radiometry across the remaining parts of the scene.
The assumption here is that the photo-textured scene was captured under diffuse daylight

so that no illumination variation exists. This assumption limits the general applicability of the approach, which might be overcome by using *intrinsic image* decomposition which decomposes an image into its reflectance and shading components. In this framework [61], the decomposition of an image is expressed as $I(x) = S(x)R(x) + C(x)$, were $I(x)$ is the image pixel value it location x, $S(x)$ is shading, $R(x)$ is the albedo, and $C(x)$ is a specular term. The idea is, instead of using the photo-textured scene (calculated based on $I(x)$) to estimate the affinities between scene points, and to use the reflectance $R(x)$ as affinity information.

I presented a novel scanning principle that results in a geo-referenced Euclidean reconstruction with radiometric surface properties beyond photo-texture. The presented approach was used during the 3D Pitoti project to scan and reconstruct rock-art in its landscape context.

Bibliography

[1] 3D COFORM project. http://www.3d-coform.eu.

[2] 3D-PITOTI project. http://www.3d-pitoti.eu/.

[3] Hpatches: Homography-patches dataset. https://github.com/featw/hpatches. Accessed: 2016-10-20.

[4] Multi-View Environment. http://www.gcc.tu-darmstadt.de/home/proj/mve/. Accessed: 2017-06-17.

[5] open Multiple View Geometry (openMVG). http://imagine.enpc.fr/~moulonp/openMVG/. Accessed: 2017-06-17.

[6] ROVINA proj. http://www.rovina-project.eu.

[7] SiftGPU. https://github.com/pitzer/SiftGPU. Accessed: 2017-06-15.

[8] The Digital Michelangelo project. https://graphics.stanford.edu/projects/mich/.

[9] The Parthenon project. http://gl.ict.usc.edu/films/parthenon/.

[10] A. Abadpour and S. Kasaei. An efficient pca-based color transfer method. *J. Vis. Comun. Image Represent.*, 18(1):15–34, Feb. 2007.

[11] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. Seitz, and R. Szeliski. Building Rome in a day. *Communications of the ACM*, 54(10):105–111, 2011.

[12] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org.

[13] C. Alexander, A. Pinz, and C. Reinbacher. Multi-scale 3d rock-art recording. *Digital Applications in Archaeology and Cultural Heritage*, 2(2):181–195, 2015.

[14] E. Anati. *Evoluzione e stile nell'arte rupestre camuna.* Archivi (Centro Camuno di Studi Preistorici). Edizioni del centro, 1975.

[15] A. Arca. *Mount Bego and Valcamonica, most ancient engraving phases comparison. From Neolithic to Early Bronze Age, parallels and differences between Marvegie (marvels) and Pitoti (puppets) of the two main alpine rock art poles.* Rivista di scienze preistoriche. Istituto Italiano di Preistoria e Protostoria, Firenze, ITALIE , 2009.

[16] M. Ashikmin, S. Premože, and P. Shirley. A microfacet-based brdf generator. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 65–74, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[17] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. In *Proceedings of the 11th European Conference on Computer Vision Conference on Computer Vision: Part III*, ECCV'10, pages 29–43, Berlin, Heidelberg, 2010. Springer-Verlag.

[18] M. Ben-Ezra, J. Wang, B. Wilburn, X. Li, and L. Ma. An led-only brdf measurement device. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[19] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[20] L. Bottou. Online algorithms and stochastic approximations. In D. Saad, editor, *Online Learning and Neural Networks*. Cambridge University Press, Cambridge, UK, 1998. revised, oct 2012.

[21] J. Y. Bouguet. Camera calibration toolbox for Matlab, 2008.

[22] G. Bradski. OpenCV. *Dr. Dobb's Journal of Software Tools*, 2000.

[23] S. Cao and N. Snavely. Learning to match images in large-scale collections. In *Computer Vision - ECCV 2012. Workshops and Demonstrations - Florence, Italy, October 7-13, 2012, Proceedings, Part I*, pages 259–270, 2012.

[24] M. Chandraker. What camera motion reveals about shape with unknown brdf. In *Proc. CVPR*, June 2014.

[25] M. Chandraker, J. Bai, and R. Ramamoorthii. A theory of photometric reconstruction for unknown isotropic reflectances. In *Proc. CVPR*, 2011.

[26] M. Chandraker, D. Reddy, Y. Wang, and R. Ramamoorthi. What object motion reveals about shape with unknown brdf and lighting. In *Proc. CVPR*, 2013.

[27] C. Chippindale and F. Baker. *PITOTI digital rock-art from prehistoric Europe: heritage, film, archaeology.* Skira Editore S.p.A., 2012.

[28] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu. Color harmonization. *ACM Trans. Graph.*, 25(3):624–630, July 2006.

[29] R. L. Cook and K. E. Torrance. A reflectance model for computer graphics. *ACM Trans. Graph.*, 1(1):7–24, Jan. 1982.

[30] K. J. Dana. Brdf/btf measurement device. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 460–466 vol.2, 2001.

[31] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, pages 189–198, New York, NY, USA, 1998. ACM.

[32] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 145–156, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[33] P. Debevec, C. Tchou, A. Gardner, T. Hawkins, C. Poullis, J. Stumpfel, A. Jones, N. Yun, P. Einarsson, T. Lundgren, M. Fajardo, and P. Martinez. Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical Report ICT-TR-06.2004, Univ. of Southern Calif. Creative Technologies Graphics Lab., 2004.

[34] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[35] M. Dellepiane, M. Callieri, F. Ponchio, and R. Scopigno. Mapping highly detailed colour information on extremely dense 3d models: The case of david's restoration. *Computer Graphics Forum*, 27(8):2178–2187, 2008.

[36] M. Dellepiane, R. Marroquim, M. Callieri, P. Cignoni, and R. Scopigno. Flow-based local optimization for image-to-geometry projection. *IEEE Transactions on Visualization and Computer Graphics*, 18(3):463–474, Mar. 2012.

[37] H. S. Faridul, T. Pouli, C. Chamaret, J. Stauder, A. Tremeau, and E. Reinhard. A Survey of Color Mapping and its Applications. In S. Lefebvre and M. Spagnuolo, editors, *Eurographics 2014 - State of the Art Reports*. The Eurographics Association, 2014.

[38] O. Faugeras, Q.-T. Luong, and T. Papadopoulou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, Cambridge, MA, USA, 2001.

[39] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proceedings of the Second European Conference on Computer Vision*, ECCV '92, pages 563–578, London, UK, UK, 1992. Springer-Verlag.

[40] O. D. Faugeras, Q. T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Computer Vision — ECCV'92: Second European Conference on Computer Vision Santa Margherita Ligure, Italy, May 19–22,*

1992 Proceedings, pages 321–334, Berlin, Heidelberg, 1992. Springer Berlin Heidelberg.

[41] S. Ferradans, N. Papadakis, G. Peyré, and J.-F. Aujol. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882, 2014.

[42] J. Filip and R. Vávra. Template-based sampling of anisotropic brdfs. *Computer Graphics Forum (Proceedings of Pacific Graphics 2014)*, 2014.

[43] J. Filip, R. Vávra, M. Haindl, P. id, M. Krupika, and V. Havran. Brdf slices: Accurate adaptive anisotropic appearance acquisition. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1468–1473, June 2013.

[44] J. Filip, R. Vávra, M. Haindl, P. id, M. Krupika, and V. Havran. Brdf slices: Accurate adaptive anisotropic appearance acquisition. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1468–1473, June 2013.

[45] S. Finsterwalder. Die geometrischen Grundlagen der Photogrammetrie. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 6:1–42, 1897.

[46] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[47] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of vision*, 3(5):3–3, 2003.

[48] R. Fletcher. *Practical Methods of Optimization*. Wiley-Interscience, New York, NY, USA, 2^{nd} edition, 1987.

[49] W. Förstner. A Feature Based Correspondence Algorithm for Image Matching. *Int. Arch. of Photogrammetry and Remote Sensing*, 26(3):150–166, 1986.

[50] W. Förstner. A framework for low level feature extraction. In *Proceedings of the Third European Conference on Computer Vision (Vol. II)*, ECCV '94, pages 383–394, Secaucus, NJ, USA, 1994. Springer-Verlag New York, Inc.

[51] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, pages 281–305, 1987.

[52] Y. Furukawa and C. Hernández. Multi-view stereo: A tutorial. *Found. Trends. Comput. Graph. Vis.*, 9(1-2):1–148, June 2015.

[53] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376, 2010.

[54] S. Gauglitz, T. Höllerer, and M. Turk. Evaluation of interest point detectors and feature descriptors for visual tracking. *Int. J. Comput. Vision*, 94(3):335–360, Sept. 2011.

[55] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.

[56] S. Georgoulis, M. Proesmans, and L. Van Gool. Tackling shapes and brdfs head-on. In *3D Vision (3DV), 2014 2nd International Conference on*, volume 1, pages 267–274. IEEE, 2014.

[57] R. Gherardi, M. Farenzena, and A. Fusiello. Improving the efficiency of hierarchical structure-and-motion. In *Proc. CVPR*, 2010.

[58] A. Ghosh, S. Achutha, W. Heidrich, and M. O'Toole. Brdf acquisition with basis illumination. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Oct 2007.

[59] M. Goesele, H. P. A. Lensch, J. Lang, C. Fuchs, and H.-P. Seidel. Disco: Acquisition of translucent objects. *ACM Trans. Graph.*, 23(3):835–844, Aug. 2004.

[60] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[61] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2335–2342, Sept 2009.

[62] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2011)*, 30(4):70:1–70:9, 2011.

[63] M. J. Hannah. *Computer Matching of Areas in Stereo Images.* PhD thesis, Stanford, CA, USA, 1974. AAI7427032.

[64] C. Harris and M. Stephens. A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[65] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 761–764. IEEE, 1992.

[66] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proceedings of the Second European Conference on Computer Vision*, ECCV '92, pages 579–587, London, UK, UK, 1992. Springer-Verlag.

[67] R. I. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(10):1036–1041, Oct. 1994.

[68] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(6):580–593, June 1997.

[69] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[70] G. Hauck. Neue Constructionen der Perspective und Photogrammetrie (Theorie der trilinearen Verwandtschaft ebener Systeme). *Journal für die reine und angewandte Mathematik*, (95):1–35, 1883.

[71] O. Hesse. Cubische Gleichung, von welcher die Lösung des Problems der Homographie von M. Chasles abhängt. *Crelle-Borchardt'schen Journal*, (62):188, 1863.

[72] A. Heyden and K. Astrom. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 438–443, Jun 1997.

[73] T. Höll, G. Holler, and A. Pinz. A novel high accuracy 3d scanning device for rock-art sites. In *ISPRS Technical Commission V Symposium*, pages 285–291, 2014.

[74] T. Höll and A. Pinz. Cultural heritage acquisition: Geometry-based radiometry in the wild. In *3D Vision (3DV), 2015 International Conference on*, pages 389–397, Oct 2015.

[75] C. Hoppe, M. Klopschitz, M. Rumpler, A. Wendel, S. Kluckner, H. Bischof, and G. Reitmayr. Online feedback for structure-from-motion image acquisition. In *Proc. BMVC*, 2012.

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

[77] B. K. Horn. *Robot Vision*. McGraw-Hill, 1986.

[78] H. Jégou, R. Tavenard, M. Douze, and L. Amsaleg. Searching in one billion vectors: re-rank with source coding. In *ICASSP 2011 - International Conference on Acoustics, Speech and Signal Processing*, pages 861–864, Prague, Czech Republic, May 2011. IEEE.

[79] N. Kobyshev, H. Riemenschneider, and L. V. Gool. Matching features correctly through semantic understanding. In *Proceedings of the 2014 2Nd International Conference on 3D Vision - Volume 01*, 3DV '14, pages 472–479, Washington, DC, USA, 2014. IEEE Computer Society.

[80] A. Kunert, A. Kulik, S. Beck, and B. Froehlich. Photoportals: Shared references in space and time. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 1388–1399, New York, NY, USA, 2014. ACM.

[81] S. Lackner and W. Lienhart. Impact of prism type and prism orientation on the accuracy of automated total station measurements. In *3rd Joint International Symposium on Deformation Monitoring (JISDM)*, pages 1–8, Vienna, Austria, 2016.

[82] N. Lackovic, C. Crook, S. Cobb, S. Shalloe, and M. D'Cruz. Imagining technology-enhanced learning with heritage artefacts: teacher-perceived potential of 2d and 3d heritage site visualisations. *Educational Research*, 57(3):331–351, 2015.

[83] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. *ACM Trans. Graph.*, 23(3):689–694, Aug. 2004.

[84] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 131–144, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[85] T. Lindeberg. Feature detection with automatic scale selection. *Int. J. Comput. Vision*, 30(2):79–116, Nov. 1998.

[86] S. Lombardi and K. Nishino. Reflectance and natural illumination from a single image. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VI*, ECCV'12, pages 582–595, Berlin, Heidelberg, 2012. Springer-Verlag.

[87] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828):133–135, 1981.

[88] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, page 131 Vol. 1, 1999.

[89] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[90] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[91] C. Lundquist, Z. Sjanic, and F. Gustafsson. *Statistical Sensor Fusion: Exercises*. Studentlitteratur AB, Sweden, 2015.

[92] S. E. Lynch, M. S. Drew, and G. D. Finlayson. Colour constancy from both sides of the shadow edge. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 899–906, Dec 2013.

[93] T. Malzbender, D. Gelb, and H. Wolters. Polynomial texture maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 519–528, New York, NY, USA, 2001. ACM.

[94] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. BMVC*, pages 36.1–36.10, 2002. doi:10.5244/C.16.36.

[95] J. Matas, K. Mikolajczyk, T. Tuytelaars, and A. Vedaldi. ECCV2016 Workshop: Local features: State of the art, open problems and performance evaluation. `http://www.iis.ee.ic.ac.uk/ComputerVision/DescrWorkshop/index.html`. Accessed: 2016-10-20.

[96] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. Graph.*, 22(3):759–769, July 2003.

[97] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *Int. J. Comput. Vision*, 8(2):123–151, Aug. 1992.

[98] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Int. J. Comput. Vision*, 60(1):63–86, Oct. 2004.

[99] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, Oct. 2005.

[100] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, Nov. 2005.

[101] R. Mohr, B. Boufama, and P. Brand. Accurate projective reconstruction. In *Proceedings of the Second Joint European - US Workshop on Applications of Invariance in Computer Vision*, pages 257–276, London, UK, UK, 1994. Springer-Verlag.

[102] H. P. Morevec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'77, pages 584–584, San Francisco, CA, USA, 1977. Morgan Kaufmann Publishers Inc.

[103] C. Mostegel, G. Poier, C. Reinbacher, M. Hofer, F. Fraundorfer, H. Bischof, T. Höll, G. Holler, and A. Pinz. The 3d-pitoti project with a focus on multi-scale 3d reconstruction using semi-autonomous uavs. In K. Niel, P. M. Roth, and M. Vincze, editors, *1st OAGM-ARW Joint Workshop*, 2016.

[104] C. Mostegel, M. Rumpler, F. Fraundorfer, and H. Bischof. UAV-based autonomous image acquisition with multi-view stereo quality assurance by confidence prediction. In *Computer Vision and Pattern Recognition Workshops, IEEE Conference on*, Jun 2016.

[105] M. Muja and D. G. Lowe. Scalable nearest neighbor algorithms for high dimensional data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36, 2014.

[106] A. Mustafa, H. Kim, E. Imre, and A. Hilton. Segmentation based features for wide-baseline multi-view reconstruction. In *3D Vision (3DV), 2015 International Conference on*, pages 282–290, Oct 2015.

[107] S. Nene and S. Nayar. Stereo with mirrors. In *Proc. ICCV*, 1998.

[108] A. Ngan, F. Durand, and W. Matusik. Experimental analysis of brdf models. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, EGSR '05, pages 117–126, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[109] K. Nishino. Directional statistics brdf model. In *2009 IEEE 12th International Conference on Computer Vision*, pages 476–483, Sept 2009.

[110] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 2161–2168, Washington, DC, USA, 2006. IEEE Computer Society.

[111] D. Nistér and H. Stewénius. Linear time maximally stable extremal regions. *Computer Vision–ECCV 2008*, pages 183–196, 2008.

[112] I. Nurutdinova and A. Fitzgibbon. Towards pointless structure from motion: 3d reconstruction and camera parameters from general 3d curves. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2363–2371, Dec 2015.

[113] U. S. N. B. of Standards and F. E. Nicodemus. *Geometrical considerations and nomenclature for reflectance*, volume 160. US Department of Commerce, National Bureau of Standards, 1977.

[114] T. Oskam, A. Hornung, R. W. Sumner, and M. Gross. Fast and stable color balancing for images and augmented reality. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 49–56, Oct 2012.

[115] OSRAM Opto Semiconductors GmbH, Leibnizstraße 4, D-93055 Regensburg. *OSLON SSL80 Datasheet*, 2012.

[116] G. Oxholm and K. Nishino. Multiview shape and reflectance from natural illumination. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2163–2170, June 2014.

[117] M. Pharr and G. Humphreys. *Physically Based Rendering: From Theory To Implementation*. Morgan Kaufmann, 2^{nd} edition, 2010.

[118] R. Pintus, E. Gobbetti, M. Callieri, and M. Dellepiane. Techniques for seamless color registration and mapping on dense 3d models. In N. Masini and F. Soldovieri, editors, *Sensing the Past: From artifact to historical site*, pages 355–376. Springer International Publishing, Cham, 2017.

[119] F. Pitié, A. C. Kokaram, and R. Dahyot. N-dimensional probablility density function transfer and its application to colour transfer. *IEEE International Conference on Computer Vision (ICCV)*, 2:1434–1439, 2005.

[120] F. Pitié, A. C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.*, 107(1-2):123–137, July 2007.

[121] M. Pollefeys. Visual 3d modeling from images. In *Proceedings of the Vision, Modeling, and Visualization Conference 2004 (VMV 2004), Stanford, California, USA, November 16-18, 2004*, page 3, 2004.

[122] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Sixth International Conference on Computer Vision (ICCV), 1998.*, pages 90–95, Jan 1998.

[123] T. Pouli and E. Reinhard. Progressive histogram reshaping for creative color transfer and tone reproduction. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*, NPAR '10, pages 81–90, New York, NY, USA, 2010. ACM.

[124] K. Pulli, H. Abi-Rached, T. Duchamp, L. G. Shapiro, and W. Stuetzle. Acquisition and visualization of colored 3d objects. In *Proceedings of the 14th International Conference on Pattern Recognition-Volume 1 - Volume 1*, ICPR '98, pages 11–, Washington, DC, USA, 1998. IEEE Computer Society.

[125] J. Rabin, S. Ferradans, and N. Papadakis. Adaptive color transfer with relaxed optimal transport. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4852–4856, Oct 2014.

[126] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In A. M. Bruckstein, B. M. Haar Romeny, A. M. Bronstein, and M. M. Bronstein, editors, *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29 – June 2, 2011, Revised Selected Papers*, pages 435–446. Springer Berlin Heidelberg, 2012.

[127] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, Sep 2001.

[128] E. Reinhard, E. A. Khan, A. O. Akyuz, and G. M. Johnson. *Color imaging: fundamentals and applications.* A.K. Peters, 2008.

[129] F. Romeiro, Y. Vasilyev, and T. Zickler. Passive reflectometry. In *Proc. ECCV*, 2008.

[130] F. Romeiro and T. Zickler. Blind reflectometry. In *Proceedings of the 11th European Conference on Computer Vision: Part I*, ECCV'10, pages 45–58, Berlin, Heidelberg, 2010. Springer-Verlag.

[131] A. S. W. Ron O. Dror, Edward H. Adelson. Estimating surface reflectance properties from images under unknown illumination. *Proc.SPIE*, 4299:4299 – 4299 – 12, 2001.

[132] M. Rothermel, K. Wenzel, D. Fritsch, and N. Haala. SURE: photogrammetric surface reconstruction from imagery. In *Proc. LC3D Workshop*, 2012.

[133] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Sixth International Conference on Computer Vision (ICCV), 1998.*, pages 59–66, Jan 1998.

[134] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao. Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A*, 15:2036–2045, 1998.

[135] M. Schaich, P. Jahnke, M. Rahrig, S. Knechtel, and M. Holzapfel. 3D-PITOTI D3.1 Ground-truth dataset. Technical report, 2013.

[136] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *Int. J. Comput. Vision*, 37(2):151–172, June 2000.

[137] C. Schwartz, R. Sarlette, M. Weinmann, M. Rump, and R. Klein. Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the university of bonn. *Sensors*, 14(5), Apr. 2014.

[138] M. Seidl, E. Wieser, and C. Alexander. Automated classification of petroglyphs. *Digital Applications in Archaeology and Cultural Heritage*, 2(2–3):196 – 212, 2015. Digital imaging techniques for the study of prehistoric rock art.

[139] V. Sequeira and J. G. M. Goncalves. 3d reality modelling: photo-realistic 3d models of real world scenes. In *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*, pages 776–783, 2002.

[140] L. Shen, T. W. Chua, and K. Leman. Shadow optimization from structured deep edge detection. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2067–2074, June 2015.

[141] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 593–600, Jun 1994.

[142] S. Shirdhonkar and D. W. Jacobs. Approximate earth mover's distance in linear time. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 – 8, 2008/06// 2008.

[143] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM Trans.Graph.*, 25(3):835–846, 2006.

[144] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4):66:1–66:11, July 2015.

[145] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision.* Brooks/Cole Publishing Company, 1999.

[146] M. Steger. Constrained BRDF estimation using a stereo scanner and controlled illumination. Technical report, Institute of Electrical Measurement and Measurement Signal Processing, TU Graz, 2015.

[147] P. Sturm. A historical survey of geometric computer vision. In P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, and W. Kropatsch, editors, *Computer Analysis of Images and Patterns: 14th International Conference, CAIP 2011, Seville, Spain, August 29-31, 2011, Proceedings, Part I*, pages 1–8. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[148] R. Sturm. Das Problem der Projektivität und seine Anwendung auf die Flächen zweiten Grades. *Mathematische Annalen*, (1):553–573, 1869.

[149] C. Tchou, J. Stumpfel, P. Einarsson, M. Fajardo, and P. Debevec. Unlighting the parthenon. In *ACM SIGGRAPH 2004 Sketches*, SIGGRAPH '04, pages 80–, New York, NY, USA, 2004. ACM.

[150] C. Tomasi and T. Kanade. Detection and tracking of point features. 1991.

[151] P. H. S. Torr, P. A. Beardsley, and D. W. Murray. Robust vision. In *Proc. BMVC*, pages 14.1–14.10, 1994. doi:10.5244/C.8.14.

[152] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ICCV '99, pages 298–372, London, UK, UK, 2000. Springer-Verlag.

[153] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(3):177–280, July 2008.

[154] S. Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London B: Biological Sciences*, 203(1153):405–426, 1979.

[155] C. Villani. *Topics in optimal transportation.* Graduate studies in mathematics. American Mathematical Society, cop., 2003.

[156] P. Viola and W. M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, Sep 1997.

[157] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[158] G. J. Ward. Measuring and modeling anisotropic reflection. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '92, pages 265–272, New York, NY, USA, 1992. ACM.

[159] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color to greyscale images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 277–280. ACM, 2002.

[160] C.-L. Wen, C.-H. Hsieh, B.-Y. Chen, and M. Ouhyoung. Example-based multiple local color transfer by strokes. *Computer Graphics Forum*, 27(7):1765–1772, 2008.

[161] T. Werner, T. Pajdla, and V. Hlaváč. Oriented projective reconstruction. In M. Gengler, M. Prinz, and E. Schuster, editors, *Pattern Recognition and Medical Computer Vision: 22-nd Workshop of the Austrian Association for Pattern Recognition (ÖAGM/IAPR)*, pages 245–254, Wien, Austria, May 14–15 1998. Österreichische Computer Gesselschaft.

[162] T. Weyrich, J. Lawrence, H. Lensch, S. Rusinkiewicz, and T. Zickler. Principles of appearance acquisition and representation. In *ACM SIGGRAPH 2008 Classes*, SIGGRAPH '08, pages 80:1–80:119, New York, NY, USA, 2008. ACM.

[163] P. Wieschollek, O. Wang, A. Sorkine-Hornung, and H. Lensch. Efficient Large-scale Approximate Nearest Neighbor Search on the GPU. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2027–2035, 2016.

[164] S. Wolf. Color correction matrix for digital still and video imaging systems. Technical Report TM-04-406, National Telecommunic. and Inf. Admin. [Washington D.C.], 2003.

[165] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980.

[166] C. Wu. Towards linear-time incremental structure from motion. In *2013 International Conference on 3D Vision - 3DV 2013*, pages 127–134, June 2013.

[167] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz. Multicore bundle adjustment. In *CVPR 2011*, pages 3057–3064, June 2011.

[168] F. Wu, W. Dong, Y. Kong, X. Mei, J.-C. Paul, and X. Zhang. Content-based colour transfer. *Computer Graphics Forum*, 32(1):190–203, 2013.

[169] G. Wyszecki and W. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley Series in Pure and Applied Optics. Wiley, 2000.

[170] J. Xia. Saliency-guided color transfer between images. In *International Symposium on Visual Computing*, pages 468–475. Springer, 2013.

[171] X. Xiao and L. Ma. Gradient-preserving color transfer. *Computer Graphics Forum*, 28(7):1879–1886, 2009.

[172] K. Yamamoto, T. Yendo, T. Fujii, M. Tanimoto, and D. Suter. Color correction for multi-camera system by using correspondences. In *ACM SIGGRAPH 2006 Research Posters*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.

[173] Y. Yang, H. Zhao, L. You, R. Tu, X. Wu, and X. Jin. Semantic portrait color transfer with internet images. *Multimedia Tools Appl.*, 76(1):523–541, Jan. 2017.

[174] M. Zeppelzauer, G. Poier, M. Seidl, C. Reinbacher, S. Schulter, C. Breiteneder, and H. Bischof. Interactive 3d segmentation of rock-art by enhanced depth maps and gradient preserving regularization. *J. Comput. Cult. Herit.*, 9(4):19:1–19:30, Sept. 2016.

[175] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *Int. J. Comput. Vision*, 27(2):161–195, Apr. 1998.

[176] Z. Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, Nov. 2000.

[177] Z. Zhang. Camera calibration. In G. Medioni and S. Kang, editors, *Emerging Topics in Computer Vision*, chapter 2, pages 4–43. Prentice Hall Professional Technical Reference, 2004.

[178] Z. Zhou, G. Chen, Y. Dong, D. Wipf, Y. Yu, J. Snyder, and X. Tong. Sparse-as-possible svbrdf acquisition. *ACM Trans. Graph.*, 35(6):189:1–189:12, Nov. 2016.

[179] X. Zhu and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical report, 2002.

[180] T. Zickler, R. Ramamoorthi, S. Enrique, and P. N. Belhumeur. Reflectance sharing: predicting appearance from a sparse set of images of a known shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1287–1302, Aug 2006.